Investigation of non-canonical enzyme activities associated with $tRNA^{His}$ in eukaryotes

Senior Honors Thesis

Presented in partial fulfillment of the requirements for graduation with honors research distinction in Biochemistry in the College of Engineering at The Ohio State University

By

Michael Wolfe

The Ohio State University

May 2014

Project Advisor: Dr. Jane E. Jackman, Associate Professor, Department of Biochemistry and Chemistry

Abbreviations

tRNA - Transfer ribonucleic acid

$tRNA^{His}$- Transfer ribonucleic acid histidine

Thg1 - $tRNA^{His}$ guanylyltransferase

HsaThg1 - $tRNA^{His}$ guanylyltransferase from *Homo sapiens*

SceThg1- $tRNA^{His}$ guanylyltransferase from *Sacchromyces cerevisiae*

HisRS - histidine aminoacyl tRNA synthetase

ts - temperature sensitive

dNTP- deoxy nucleotide triphosphate

AMP - adenosine monophosphate

^{32}P - phosphorus-32

ATP- adenosine triphosphate

GTP- guanosine triphosphate

EDTA - ethylenediaminetetraacetic acid

RNase A - ribonuclease A

CIP - calf intestine phosphatase

TLC - thin layer chromatography

FPLC - fast protein liquid chromatography

V_e- elution volume

SD - synthetic dextrose

SGal - synthetic galactose

FOA - 5-fluoroorotic acid

MW - molecular weight

WT - wild type

LIC - ligation independent cloning

TTP- thymidine triphosphate

Abstract:

Transfer RNA are the most heavily modified nucleic acids in the cell. These complex molecules present the difficult challenge of retaining almost identical secondary and tertiary structure while still maintaining unique identity elements such that individual tRNA species are aminoacylated with the correct amino acid by their cognate amino-acyl tRNA synthetase. To accomplish this specificity, tRNA species have coevolved with a diverse set of enzymes that recognize and modify each tRNA to the mature form used by the ribosomal machinery. The interplay between $tRNA^{His}$guanylyltransferase (Thg1) and histidine aminoacyl tRNA synthetase (HisRS) in the context of $tRNA^{His}$ maturation represent a fascinating system with which to study these questions related to tRNA identity.

Thg1 is part of a family of enzymes found in all three domains of life, members of which catalyze an unprecedented 3'-5' nucleotide addition reaction to post-transcriptionally add a critical G_{-1} residue to the 5' end of $tRNA^{His}$ that is a necessary identity element for the HisRS in organisms from all three domains of life. Structural similarities between canonical DNA/RNA polymerases and eukaryotic Thg1 suggest unexpected connections between these enzyme activities. Recent structural and transient kinetic analysis of human Thg1 revealed several important mechanistic features, including the roles of several highly conserved residues in the chemistry of 3'-5' addition. However, many mechanistic questions, particularly related to the ability of Thg1 to bind to its tRNA substrates, remain unsolved. Since tRNA is a relatively large molecule, similar in overall molecular weight to the mass of a single Thg1 monomer, and Thg1 has been observed to exist as a tetramer, it was hypothesized that the multimeric form of the enzyme is critical for recognition of the large tRNA substrate. For this reason, several highly conserved residues were investigated that are predicted to be involved in Thg1 multimerization. Previously, variants of human Thg1 believed to disrupt the dimer-dimer interface of the enzyme based on their positions in the crystal structure had been characterized in vitro. The first aim of this project was to test the function of these particular variants in vivo in yeast. This objective was accomplished

using a yeast complementation assay to test whether each variant was able to support growth in the absence of the wild type *THG1* gene. The in vivo results were found to largely mirror the previous in vitro characterization, supporting the predicted role of these particular residues in maintaining the dimer-dimer interface. The second aim of this project was to identify novel residues involved in different steps of the 3'-5' addition reaction catalyzed by Thg1. Previously, a collection of temperature-sensitive yeast Thg1 variants were identified through random mutagenesis of the yeast *THG1* gene; interestingly these variants were observed to form distinct complementation groups suggestive of possible complementary functions in the context of the wild-type enzyme. Individual variants were expressed, purified and characterized in vitro using a G_{-1}-addition activity assay. Although most variants led to strong catalytic defects consistent with severe folding defects associated with the alterations, a new interaction involving residues from the extreme N- and C-termini of Thg1 was identified through this work. The third and final aim of this project was to identify the molecular determinants for HisRS specificity in canonical and non-canonical HisRS enzymes. Investigations of diverse eukaryotes, such as *Trypanosoma brucei* and *Acanthamoeba castellanii,* revealed that the nuclear-encoded $tRNA^{His}$ in these species lacks the characteristic G_{-1} residue, and the native HisRS from these species efficiently aminoacylates $tRNA^{His}$ independent of this critical identity element. To determine the molecular basis for this difference, residues were exchanged between G_{-1}-dependent and G_{-1}-independent HisRS and the variant enzymes were tested for aminoacylation activity. This led to the identification of residue M160 in *S. cerevisiae* as a potential molecular determinant for G_{-1} specificity in canonical HisRS enzymes. Taken together, this study has widened understanding of the molecular details of interaction between enzymes that are crucial for proper $tRNA^{His}$ maturation and their substrates, fundamentally advancing our understanding of this system that is critical for life.

Introduction:

Translation, the ubiquitous process of decoding the nucleotide bases contained within mRNA to the amino acids that make up functional proteins in all known life, requires the coordination of several highly conserved molecules. Central to this process is transfer RNA (tRNA), a unique class of ribonucleic acids whose members differ in their individual anticodons that correspond to a specific amino acid but share the highly conserved secondary and tertiary structure that results in their hallmark three-dimensional L shape. This common L-shape structure has the advantage of allowing the ribosomal machinery to recognize each tRNA with a single common active site but it also presents a unique challenge for enzymes that interact specifically with tRNA molecules. In particular, to maintain fidelity of translation, each tRNA must be aminoacylated with the correct amino acid for its corresponding anticodon. This is accomplished by a selective recognition of specific tRNAs from the overall pool of similarly structured tRNAs by a cognate aminoacyl tRNA synthetase (aaRS). This recognition is made possible by using specific sequences known as identity elements on each tRNA, which often includes nucleotides in the anticodon and acceptor stem (*1, 2*). These differing nucleotides must not only set the individual tRNAs apart from each other, but must also maintain the final tertiary structure of the tRNA and therefore minimal deviations in the sequence of an individual tRNA could potentially interfere with the use of these identity elements.

When compared with other tRNAs, $tRNA^{His}$ is unique in that it contains an extra guanosine on its 5' end (G_{-1}). This G_{-1} serves as a critical identity element for the histidyl-tRNA synthetase in almost all organisms studied. Genomic sequencing revealed that this guanosine residue was not genetically encoded in eukaryotes, but added post-transcriptionally by an enzyme, although the identity of that enzyme remained a mystery for nearly 20 years (*3*). Using a biochemical genomics approach, the enzyme responsible for the nucleotide addition was discovered in *Saccharomyces cerevisiae* and named $tRNA^{His}$guanylytransferase (Thg1) (*4, 5*). Since then, homologues of Thg1 have been identified in all three

domains of life and the structure and mechanism of this Thg1 family of enzymes has been extensively studied.

The x-ray crystal structure of human Thg1 (Figure 1) revealed many important clues to the function of Thg1 (*6*). This structure, and also a more recent structure of the enzyme from the bacterium *Bacillus thuringiensis*, showed significant similarity between Thg1 and guanyl and adenyl cyclases, both important proteins in signal cascades within cells (Figure 2). Also, there is a clear similarity between Thg1 and the palm domain of many polymerases, including T7 DNA polymerase. It is noteworthy to see the similarity of the palm domain and the locations of carboxylate residues between hTHG1 and the A Family of polymerases (Figure 2B). These similarities strongly suggested that Thg1 also uses the two-metal ion mechanism seen in all canonical 5'-3' polymerases (*7*). For 5'-3' DNA/RNA polymerases, the two metal ion mechanism involves two divalent metal ions (usually magnesium) in the active site that act as coordinating structures for stabilization of charge and chelation of β, and γ phosphates on the incoming dNTP as well as activation of the attacking nucleophile during the canonical polymerase reaction (Figure 3).

All known 5'-3' polymerases catalyze attack by the polynucleotide 3'-OH on the 5'-triphosphate of an incoming NTP, which yields an added nucleotide and a pyrophosphate, thus extending the polynucleotide chain in the 5'-3' direction (Figure 4A). An identical mechanism of nucleotide addition could theoretically occur in reverse to add a nucleotide at the 5' end. In this case, attack of the 3'-OH of the incoming nucleotide on a 5'- triphosphorylated polynucleotide which would also yield an added nucleotide and a pyrophosphate, but with the net extension of the polynucleotide chain occurring in the 3'-5' direction for this mechanism (Figure 4B). Thg1 catalyzes this type of unusual nucleotide addition reaction. In vitro, the 3'-5' nucleotide addition reaction catalyzed by eukaryotic Thg1 enzymes during G_{-1} addition to $tRNA^{His}$ occurs in three distinct steps: activation of the 5' end of the tRNA by adenylylation, attack of the 3'-OH of GTP on the activated intermediate, and removal of the 5'- pyrophosphate from

the newly added guanosine residue (Figure 4C). This process yields the mature mono-phosphorylated, G_{-1}-containing, $tRNA^{His}$, AMP, and pyrophosphate (4, 5). The similarities between the reaction catalyzed by Thg1 and that of a theoretical 3'-5' polymerase raised the question of whether Thg1 may also catalyze templated addition of polynucleotides in the 3'-5' direction. Interestingly enough, templated 3'-5' polymerase activity has been associated with Thg1 enzymes, first in yeast and later with family members from all 3 domains of life (*8, 9*). This remarkable activity coupled with the previously described similarity in structure between Thg1 and canonical polymerases raises interesting questions about the mechanism, evolution, and capabilities of this enzyme.

Although much has been determined about the mechanism of the Thg1 family of enzymes, several questions still remain about the structure and molecular details of substrate recognition. It is known that the GUG anticodon is a necessary and sufficient element for recognition by yeast Thg1, yet the enzyme acts on the 5' end of $tRNA^{His}$, roughly 70 Å away from the anticodon sequence (10). The HsaThg1 crystal structure suggests that Thg1 forms a homotetramer with an asymmetric and symmetric interface, also referred to as a dimer of dimers. Based on the crystal structure, it is unlikely that a single subunit would be able to span the 70 Å distance between the anticodon and the 5' end of the tRNA suggesting that multimer formation may be needed for catalysis. Furthermore, the specific residues involved in $tRNA^{His}$ recognition are still a mystery and without a co-crystal structure with RNA these residues are difficult to predict. Finally, it was recently observed that the protozoan species *Acanthamoeba castellani* is unique in that it lacks a *THG1* gene and its HisRS does not require this additional G_{-1} residue for aminoacylation (3). Further investigation demonstrated that the eukaryotic pathogen *Trypanosoma brucei* shares this novel feature with *A. castellani* (4). This raises the question of how the HisRS in these organisms is capable of recognizing and properly amino-acylating their $tRNA^{His}$ without the crucial identity element that is seen in other eukaryotes. These three distinct questions

form the basis for this thesis project, and these investigations have yielded critical information toward understanding the complexity of proper $tRNA^{His}$ maturation.

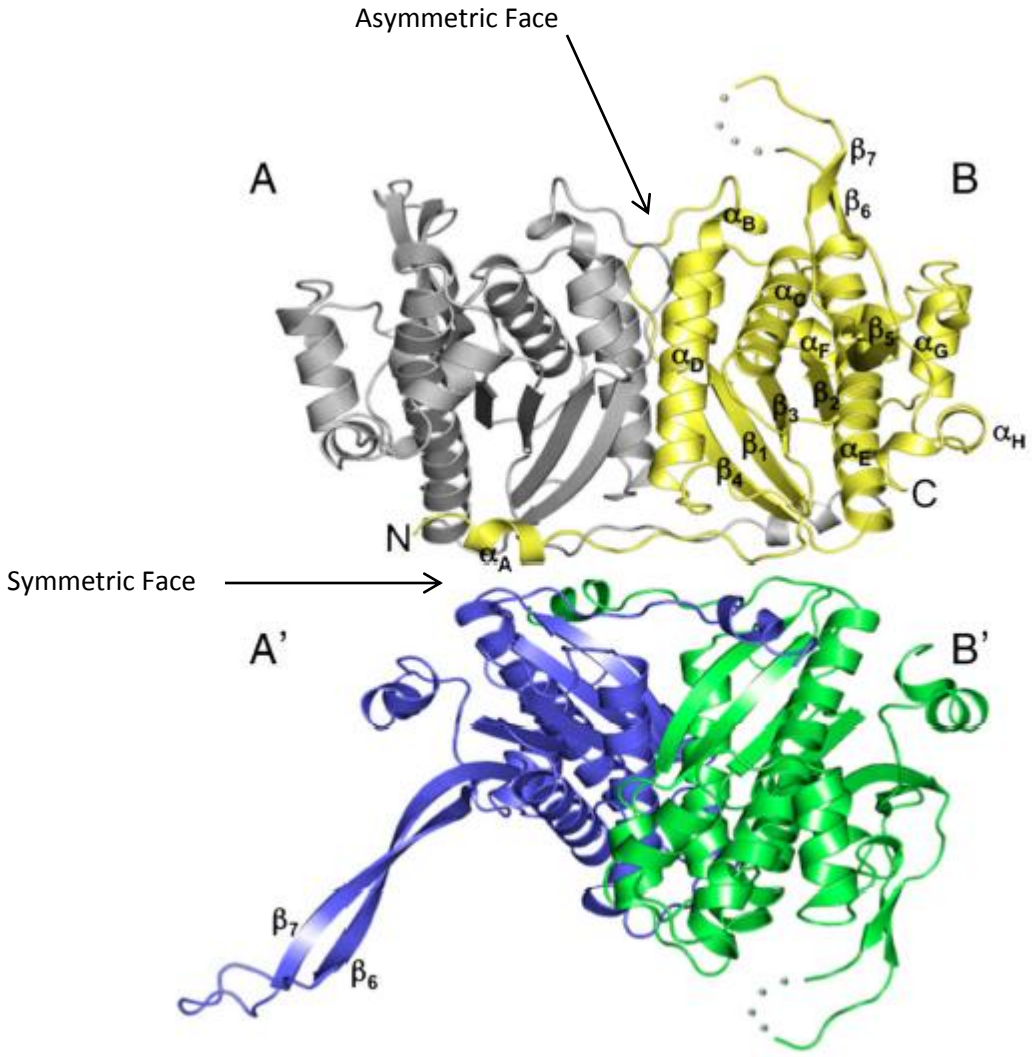


Figure 1: Thg1 crystal structure showing a homotetramer tertiary structure with a dimer of dimers architecture displaying both a symmetric and asymmetric face. Figure adapted from reference (6)

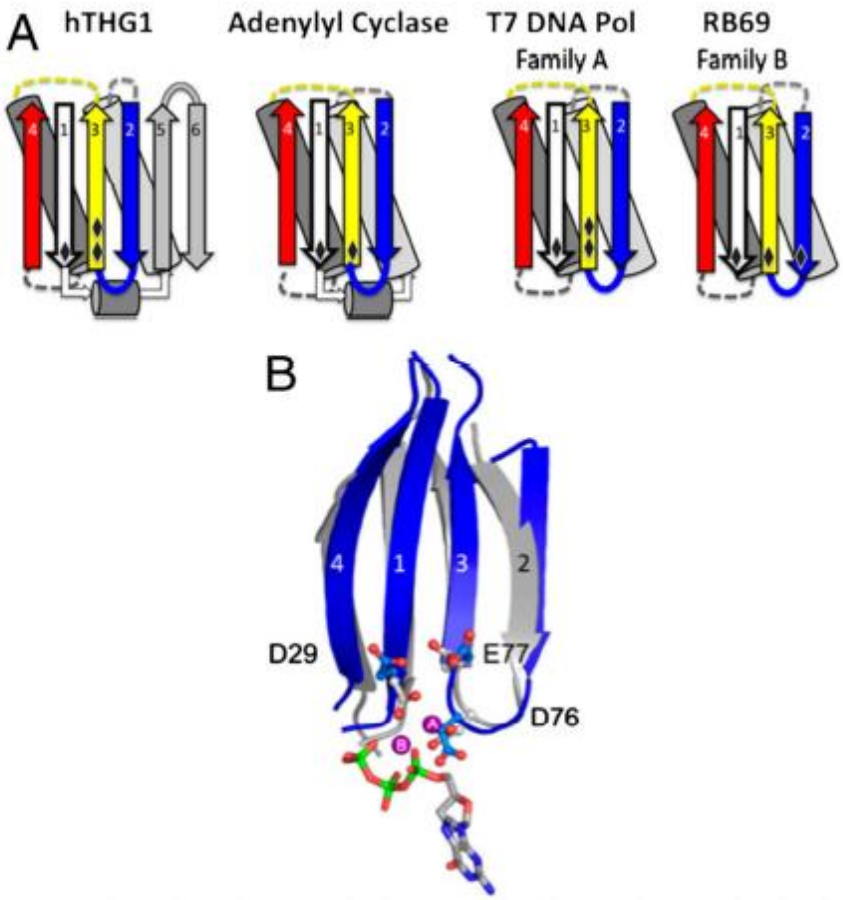


Figure 2: (A) Comparison of topology diagrams for hTHG1, adenylate cyclase ,and palm domains of T7 DNA polymerase,and RB69. (B) Superposition of hTHG1 (residues 22–135, blue) with the palm domain of T7 DNA polymerase (residues 466–486, 608–698, gray) overlays catalytic carboxylates. The incoming nucleotide and metal ions from the T7 DNA polymerase complex are shown. The β-strands are numbered for the palm domain and the locations of the carboxylate residues are displayed using black diamonds. Figure adapted from reference (6).

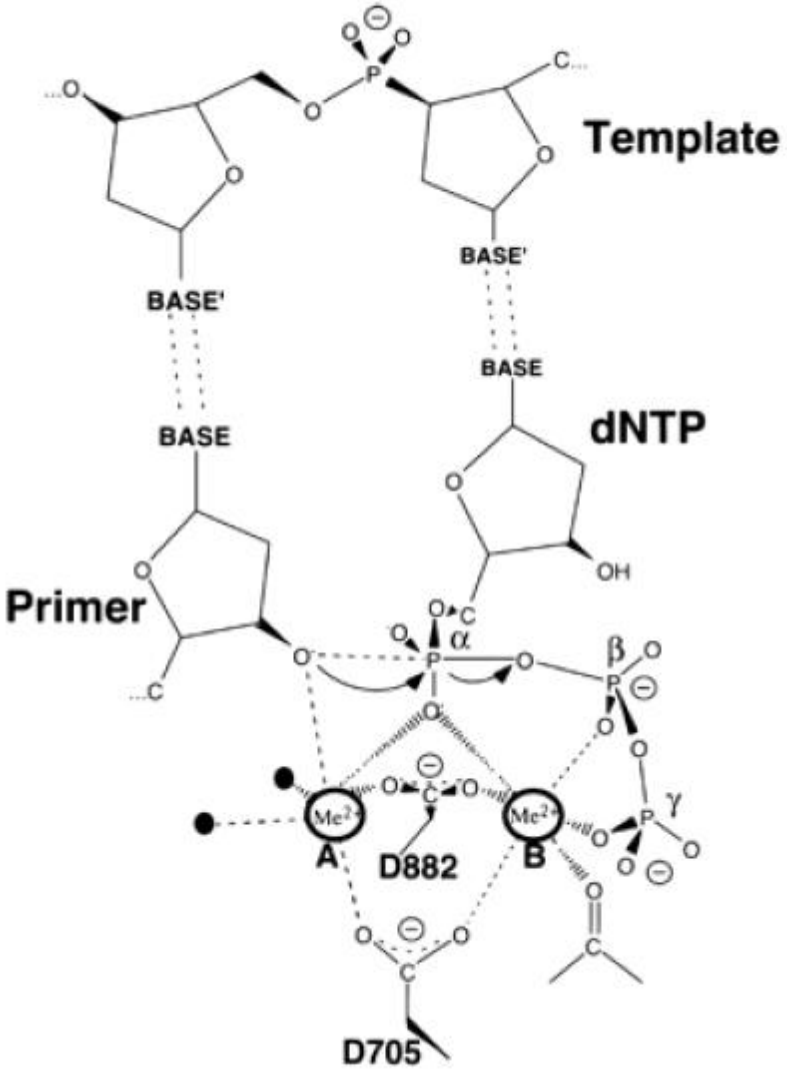


Figure 3: Two Metal Ion Mechanism for Polymerization. Clearly seen here is the two critical aspartate residues needed to coordinate to two metal ion center for catalytic activity. These two aspartates are also found in Thg1 suggesting a two metal ion mechanism of catalysis. Figure adapted from reference(7).

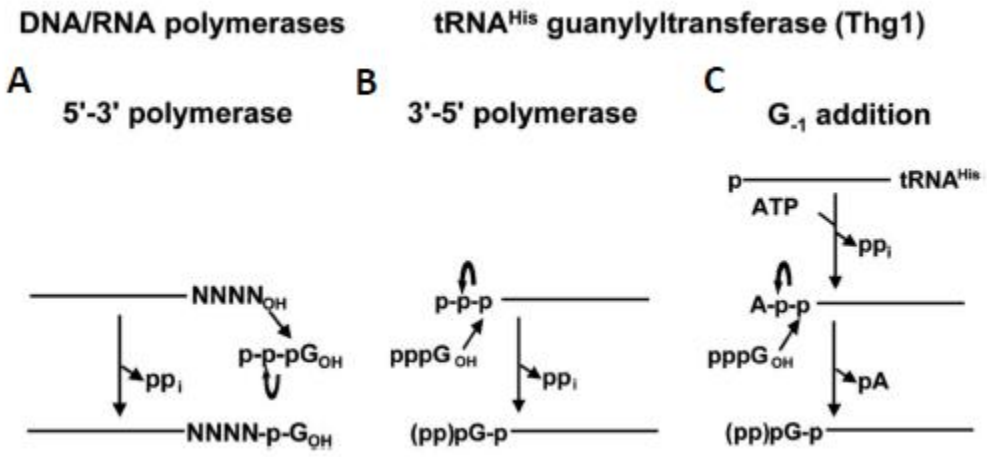


Figure 4: Comparison of potential mechanisms of polymerase activity. A. displays the 5'-3' nucleotide addition activity found in canonical polymerases. B. displays a potential mechanism for 3'-5' nucleotide addition. C. displays the mechanism of catalysis for Thg1 G_{-1} addition. Each p represents a phosphate group while pp represents a pyrophosphate group. N represents any nucleotide. Adapted from reference (5)

Chapter 1: Characterization of Thg1 quaternary structure through mutagenesis

Previously, using the crystal structure of human Thg1 (HsaThg1) as a guide, several variants were constructed that appeared to disrupt the oligomeric structure of Thg1 in vitro (6). Three universally conserved hydrogen bonding residues (T98, S102 and S106, according to the yeast numbering) were identified along two alpha helices in the asymmetric interface of the tetramer crystal structure(Figure 5). Several variants designed to affect these hydrogen-bonding interactions were created by site-directed mutagenesis and purified, including: T98A, S102A, S102A/S106A, T98Y, T98S, and T98C. Of these variants, three (T98A, S102A/S106A, T98Y) caused severe reduction in G_{-1} addition activity, while the other variants, including T98S and T98C, caused more modest defects in G_{-1} addition by Thg1 (Figure 6). Together, these data are consistent with a role for these hydrogen-bonding residues in maintaining Thg1 tetrameric structure, which is proposed to be important for catalysis.

A preliminary investigation by size-exclusion chromatography of the wild-type, T98A, S102A and S102/S106A variant showed differences in the apparent molecular weight of the major elution peak of each variant. Specifically, while wild-type and S102A variant exhibit a major peak corresponding to ~160 kDa (consistent with the crystallographically-observed tetramer) and a minor peak eluting at a volume corresponding to ~50 kDa, the T98A and S102A/S106A variants exhibited an altered elution profile, with the major peak for both enzymes corresponding to ~50 kDa (data not shown). Because the apparent molecular weights calculated from comparison to commercial standards for Thg1 are not a precise measurement of the molecular weight of each species, it is difficult to conclusively assign the 50 kDa peak to representing either the monomer or dimer conformation of Thg1. Taken together, the size-exclusion chromatography and overall activity results are consistent with a critical role for these three residues, with the enzymes that displayed strong defects in G_{-1} addition activity also exhibiting a change in apparent multimeric state by gel filtration.

Creation of vectors and yeast strains for use in in vivo characterization of structural variants

It was somewhat surprising to observe that alteration of a single hydrogen-bonding residue such as T98 could have such severe effects on the overall structure and function of Thg1, and we questioned whether the same results would be observed for these variants in the context of the physiological environment inside the yeast cell, where other contributions toward enzyme structure and stability could also play a role and might affect the interpretation of the in vitro results. To test this, we used a previously described genetic assay to assess the function of the Thg1 hydrogen-bonding residue variants in vivo in yeast.

Primers were designed for each of the variants to introduce the alteration into a parent vector consisting of the WT Hsa*THG1* gene under the control of an inducible galactose promoter on a plasmid with a *LEU2* marker [*CEN PGAL-THG1 LEU2*]; the variant HsaThg1 plasmids were created following Quick Change Site Directed mutagenesis as described in the methods section. There were difficulties creating the S106A and T98Y variants using this method so new primers were designed and these variants were created use Phusion Site-Directed Mutagenesis as described in the methods section. All variants were then confirmed by sequencing from either the OSU Plant-Microbe Genomics Facility (T98A, T98S, T98C, S102A) or Genewiz (T98Y, S106A). Although construction of the S102A/S106A double alteration was attempted repeatedly by both mutagenesis methods, we were not able to successfully obtain this variant for expression in yeast. Once confirmed, these plasmids were transformed into the *thg1* deletion yeast strain JJY20 (relevant genotype: *mat a, thg1Δ::kanR, leu2Δ, ura3Δ [CEN* P_{THG1}*-THG1 URA3]*). Since Thg1 activity is essential in yeast, the viability of this strain was maintained by the presence of a covering plasmid that encodes wild-type *THG1* under control of its native promter with a *URA3* selectable marker. Each sequenced variant HsaThg1-*LEU2* plasmid was then transformed into this strain and transformants were selected on synthetic dextrose media lacking both uracil and leucine (SD-Leu-Ura).

Yeast Complementation Assay

The purified transformants (at least two independent colonies each) were tested by replica-plating onto synthetic galactose lacking both uracil and leucine (SGal-Leu-Ura) to test if the expression of any of the variant *thg1* enzymes (under control of the Galactose-inducible promoter) caused a dominant negative phenotype. All colonies grew equally well on galactose-containing media, suggesting that the presence of the variant enzymes did not interfere with wild-type Thg1 function. Next, each strain was replica plated onto synthetic media lacking leucine and containing 5-flourourotic acid (SD-Leu+FOA and SGal-Leu+FOA) to test for complementation. Since FOA makes the presence of the *URA3* gene toxic to the yeast cell, it serves as a selective pressure to expel the wild-type *THG1* covering plasmid, leaving the variant Thg1 as the only source of Thg1 activity within the cell. If growth is observed on SGal-Leu+FOA, this indicates that the variant Thg1 catalyzes a level of G_{-1} addition activity that is sufficient for viability and therefore successfully complements the growth defect associated with deletion of *thg1*. As can be seen in Figure 7, the T98A, T98Y and empty vector (EV) do not complement yeast growth, as indicated by their lack of growth on FOA, while WT, T98S, T98C, S106A, and S102A do complement yeast growth. This result was repeated 8 times for WT, T98A,and EV; 6 times for S102A; 5 times T98S and T98C; 4 times for S106A and 3 times for T98Y. The lack of growth for all strains on SD-Leu+FOA media demonstrates that the observed complementation is indeed due to expression of the variant HsaThg1 gene, since growth in the presence of dextrose represses expression from the galactose-inducible promoter. Thus, while the lack of growth observed with T98A and T98Y variants was consistent with their observed lack of in vitro activity (and shift in apparent molecular weight of T98A Thg1), the ability of both S102A and S106A variants to fully support growth individually, combined with the inability to test the double S102A/S106A variant with this assay, left some unresolved questions about the role of these two serine residues in Thg1 multimer formation.

Size-exclusion Chromatography of Structural Variants

Therefore, to further evaluate the roles of the S102 and S106 residues in the Thg1 structure, site-directed mutagenesis was also used to create the single S106A variant that had not been created or tested in any of the earlier in vitro experiments. The S106A variant was successfully constructed in a previously-used vector for expression and purification of Thg1 variants from *E. coli*, and purified along with the wild-type enzyme, as well as the T98A and S102A/S106A variants for simultaneous comparison. The purified S016A Thg1 was obtained in sufficient yield (~0.712 mg/0.5 L cells), and with similar yield and purity to the other tested variants (data not shown).

Experiments to test the in vitro activity of the S106A variant are currently underway, but size exclusion chromatography experiments revealed substantial changes in multimeric structure associated with the S106A variant enzyme. Previous size-exclusion chromatography experiments (described above, with T98A, S102A and S102A/S106A double variant) had shown that disruption of tetramer formation correlated well with in vitro activity assays, yet these were performed under relatively high salt concentrations (500 mM NaCl) that could have affected in vitro tetramer formation. Here, the size-exclusion chromatography experiment was performed with WT HsaThg1 and variants T98A, S106A/S102A, and S106A under the buffer conditions of 50 mM sodium phosphate, 250 mM NaCl ,pH 7.2. This lower salt concentration approximates the conditions used in the in vitro assay and is therefore likely mimics the multimeric structure of the variant enzymes in the in vitro activity assays. A Superdex 200 10/300 GL column was used with an AKTA FPLC (GE Health sciences) and BioRad gel filtration standards were run in the same buffer to determine the molecular weight of the elution volumes. As seen in Figure 8, the WT hThg1 eluted with its major peak at elution volumes of ~12.5 mL (corresponding to ~166 kDa) and a minor peak at ~14 mL (corresponding to ~97kDa), which is consistent with the previous gel filtration results, particularly for the larger molecular weight peak. As observed previously, T98A Thg1 shows a significant shift in the distribution of peaks, with the largest peak

associated with the lower molecular weight species. With the S106A/S102A variant, the shift in the distribution of peaks is not as drastic as reported at the higher salt concentrations but is still consistent with disruption of tetramer formation. Interestingly, however, with the S106A variant, there was a shift in the pattern of peaks to the predominant lower molecular weight species that is similar to what is seen with the S106A/S102A variant and clearly different from that observed with wild-type Thg1.

Discussion

As predicted from the crystal structure, it appears as though the three hydrogen bonding residues along the asymmetric face of the dimer of dimers arrangement of hThg1 are important for tetramer formation. The in vitro data suggest an unequal contribution of each the residues to the overall stability of the tetramer. It is noteworthy to mention the large reduction in the activity of T98A. The tyrosine residue is thought to participate in a hydrogen bonding interaction between two adjacent monomers, thus potentially contributing to the stability of the overall tetramer structure of the enzyme. When T98 is replaced by alanine, the in vitro activity of the enzyme is significantly reduced, the predominant species migrates with a lower molecular weight on gel filtration, and the variant enzyme does not support the growth of yeast cells. The S102 and S106 residues are also observed to participate in inter-subunit hydrogen bonds in the structure, and therefore are thought to have the same function to that of the T98. However, the in vitro activity assays suggested that the roles of these residues individually are less critical. When S102 is changed to alanine it still retains much of its function and is able to support growth of yeast *thg1Δ* cells, and only when both the S102A and S106A alterations are present is activity, and mobility on gel filtration, severely affected (Figure 6, Figure 8). The isolation of the S106A alteration described in this work suggests that the roles of S102 and S106 are more complicated. Although the single S1066A variant and S102A variants do not elicit the same defects in complementation in the in vivo yeast complementation assay, the S106A variant clearly shows a defect

in multimeric structure in the size exclusion chromatography. Curiously, the S106A variant shows comparable G_{-1} addition activity to that of the WT enzyme. Additionally, the S106A/S102A variant displayed better activity at higher concentrations than previously tested, but still substantially lower activity than either of the serine residues alone (Figure 6). This data suggests that multimeric structure and subsequent activity of the serine variants is not nearly as disrupted when compared with the T98 residue and these serine residues may not be as critical for multimer formation.

A series of other replacements for T98 showed varied effects on activity suggesting T98 can be replaced by other hydrogen-bonding residues (T98S and T98C), but may be limited by overall size on electronegativity because of the defects associated with the T98Y alteration in vitro and in vivo. Strikingly, the single amino acid residue T98 also appears to have the most drastic effect on hThg1 activity in vivo as T98A and T98Y were the only variants tested that showed no complementation in yeast suggesting low enough activity as to not support life.

Taking the in vitro and in vivo data together suggests a potential threshold of activity of ~20% relative to the WT enzyme that is needed in order to provide enough activity to support normal cellular viability. The T98A variant, with 16% of the wild-type activity and T98Y with 11%, both show no complementation of the *thg1Δ* phenotype, whereas T98S at 61%, T98C at 24% and S102A at 60% activity all support growth of the same strain. This strongly suggests that tetramer formation of Thg1 is necessary for G_{-1} addition activity both in vitro and in vivo and through these experiments residues critical to the quaternary structure have been clearly identified and characterized. Unfortunately, the exact structural nature of the defective variants is not clear and since the shape of the molecule can drastically affect the elution volume in size exclusion chromatography, a better characterization of the precise multimeric structure of the lower MW peak observed in size exclusion chromatography is desired. Further, the high concentration of protein used in size-exclusion experiments does not mimic the native concentration of Thg1 in the cell and may be artificially shifting the equilibrium of the

structural state of the enzyme to the multimeric form. In order to better understand the structural basis of the variant phenotypes, more precise techniques are needed. By performing analytical ultracentrifugation, a more accurate MW could be determined for the smaller size peak and a Kd for formation of the tetramer could be determined in order to better understand the precise details of tetramer formation. Additionally, mass spectrometry would not only eliminate size effects seen in the size-exclusion chromotagraphy, but also allow for much lower concentrations of enzymes to be used in order to determine the precise multimeric state of each variant enzyme closer to the native concentration in vivo. Either one of these techniques could help shed light on the structural details of tetramer formation and give insight to similar enzymes that show this type of "dimer of dimers" interaction.

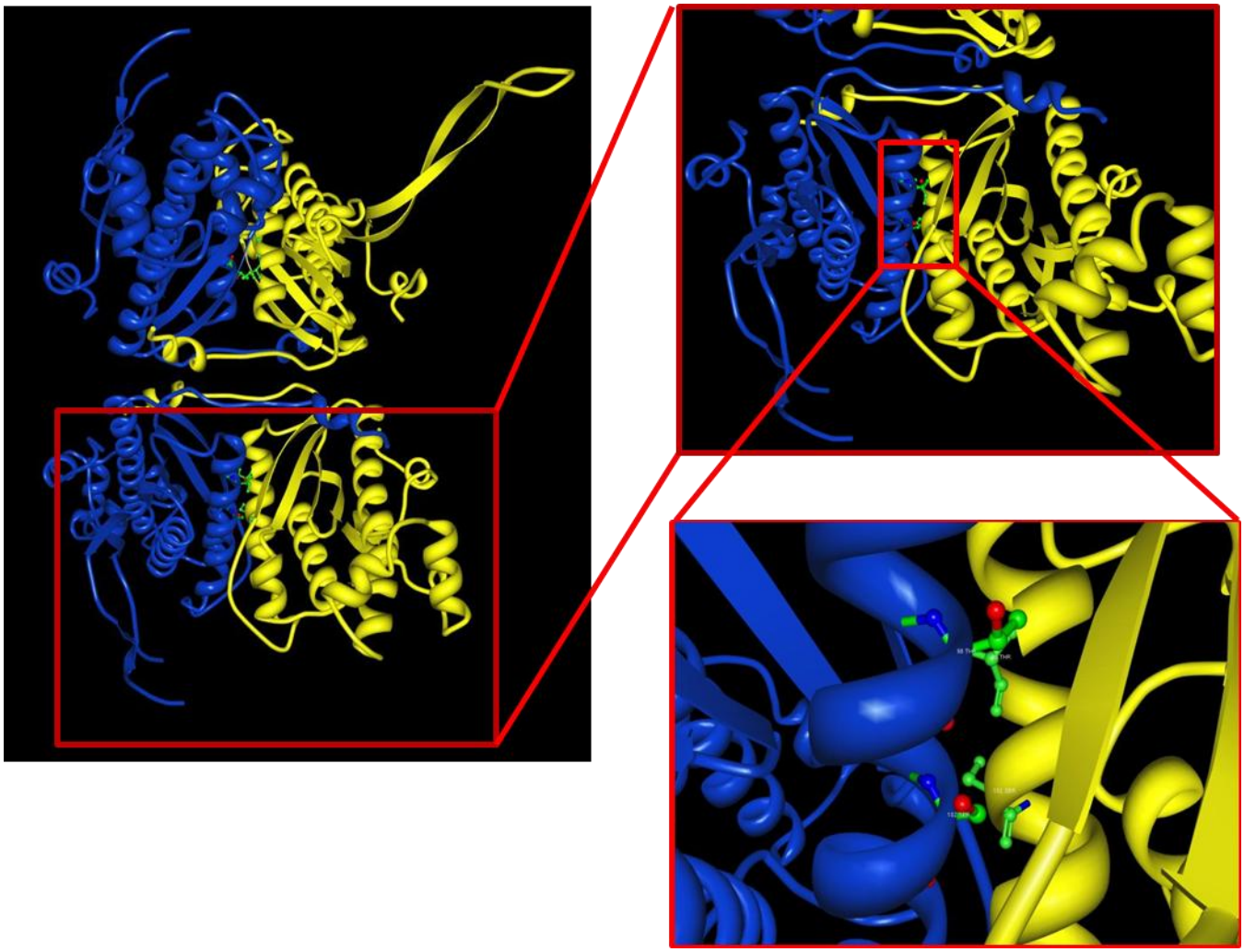
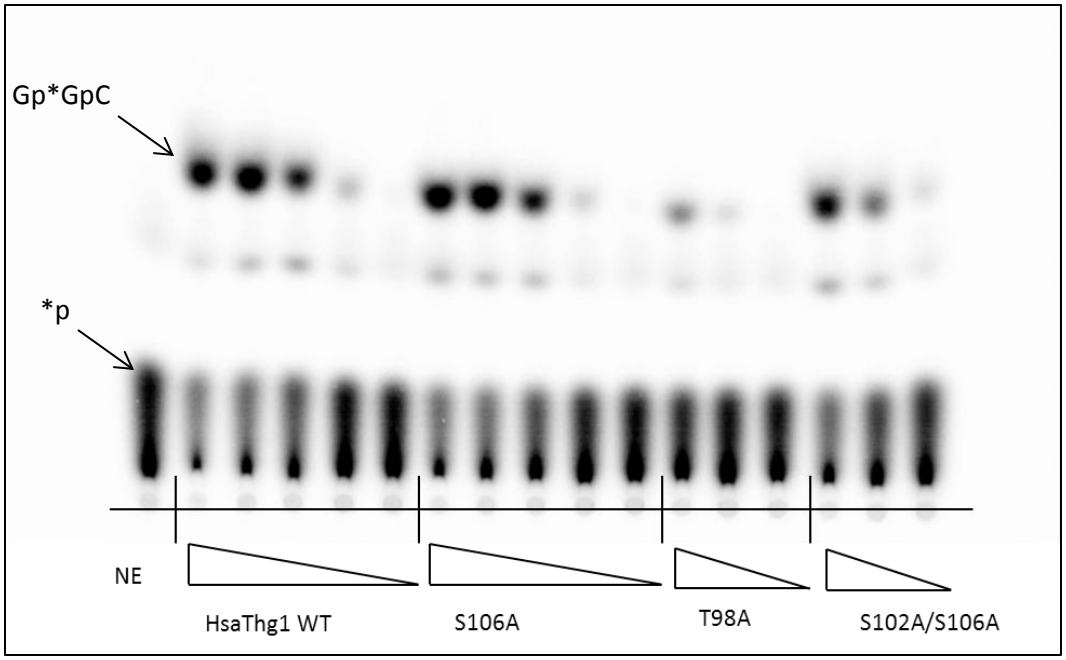
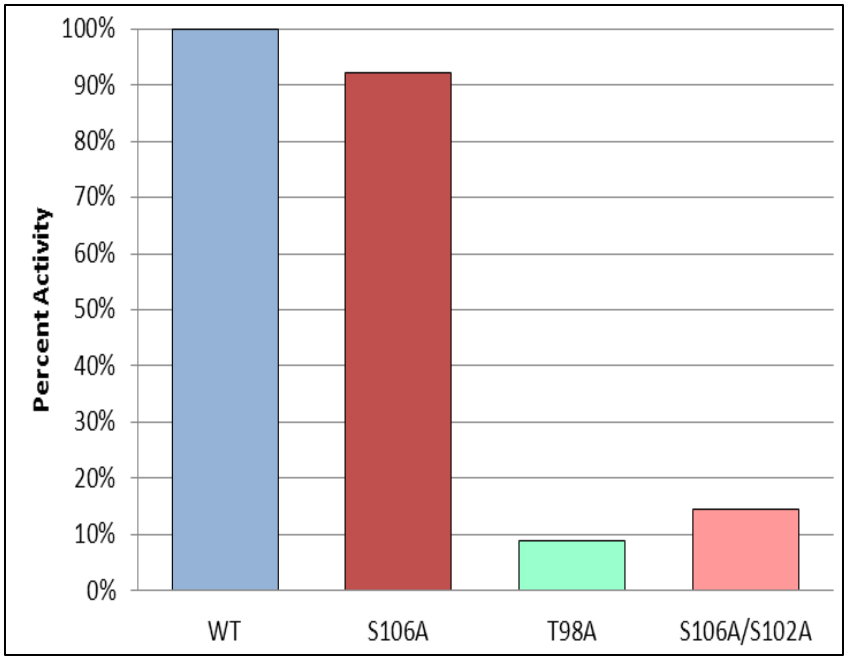


Figure 5: Crystal structure of hThg1 showing predicted hydrogen bonding residues between adjacent alpha helices in the asymmetric face of the tetramer. The residues of interest are explicitly shown from the backbone of the enzyme.





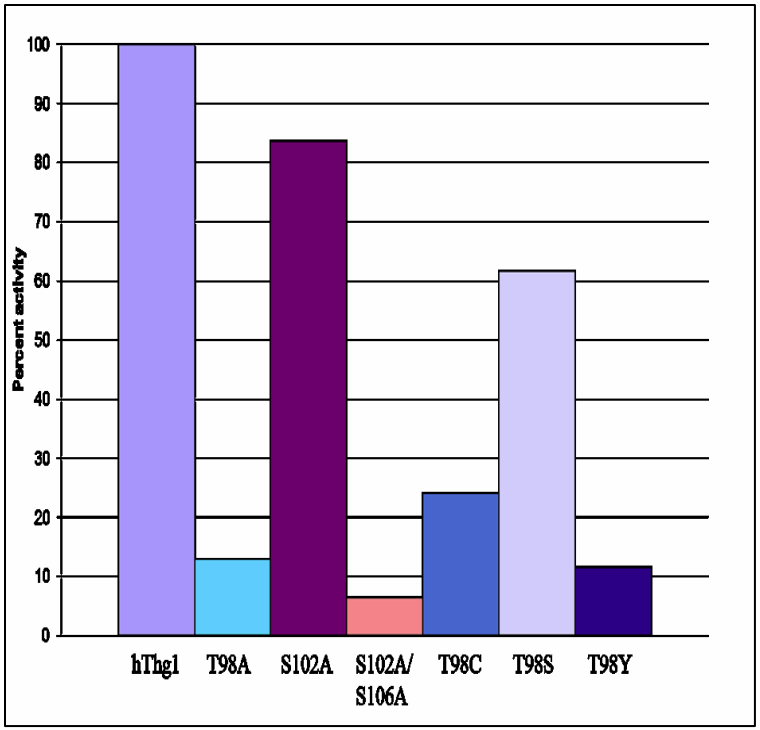


Figure 6: Variant human Thg1 activity as compared to the wild-type. The upper right panel displays a G_{-1} activity assay in which each enzyme was mixed with P^{32} labeled $tRNA^{His}$ in 0.1 mM ATP and 1 mM GTP for 2 hours, digested with RNase A and CIP, and run on a silica TLC plate in 55:35:10 n-propanol:NH_4OH:ddH_2O as the solvent. Each enzyme was tested in 5 fold serial dilutions starting at 7.5 uM. The upper right panel displays quantification of this assay based on ~100 ng of each variant run concurrently, with a wild-type substrate turnover of 29%. The bottom right panel displays previous data (Eberley and Jackman unpublished) and displays activities based on ~100 ng of each variant run concurrently, with a wild-type substrate turnover of 7%.

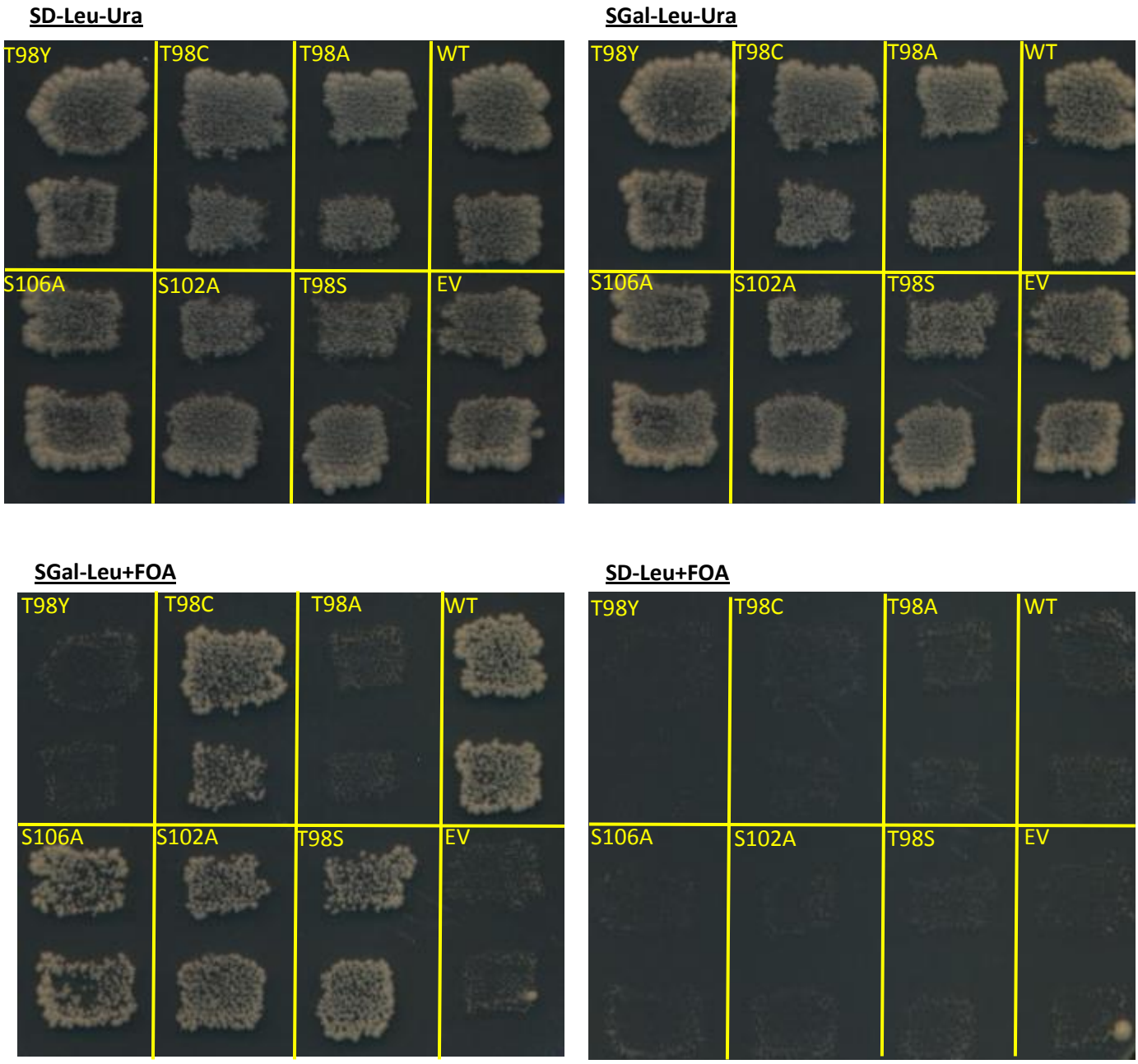
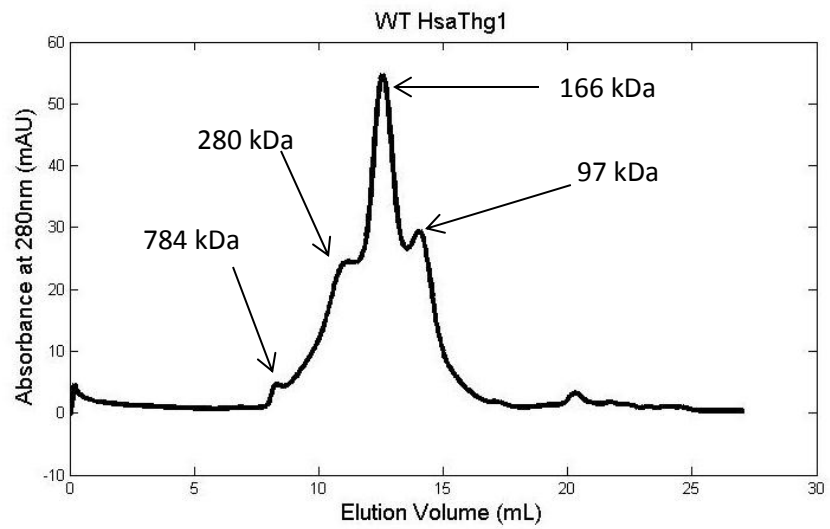
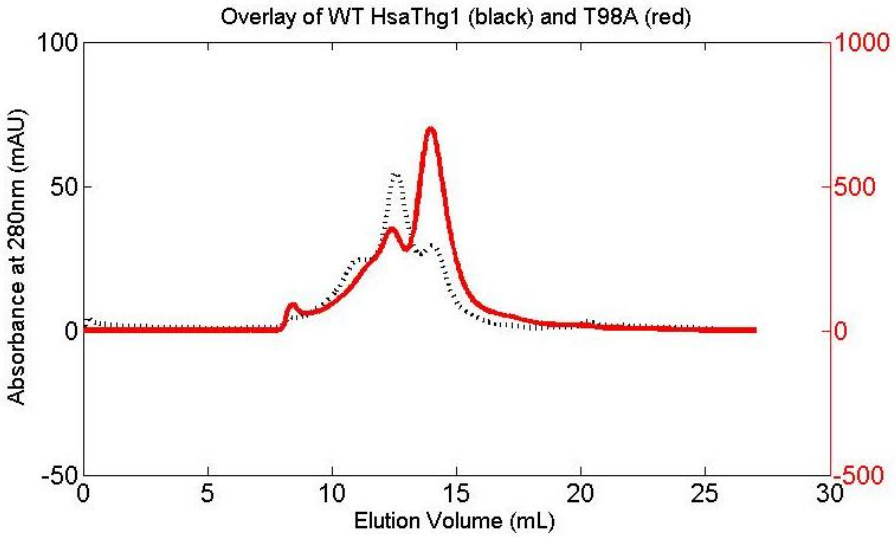
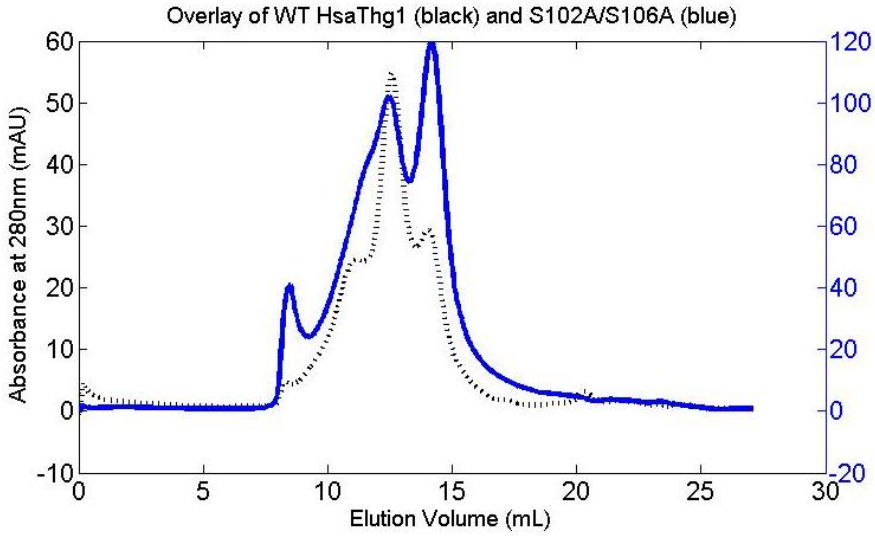


Figure 7: Growth of selected hThg1 variants in yeast at 30 °C. Plates were spotted from a master SD-Leu-Ura plate. Growth on SD-Leu-Ura suggests successful transformation of variant strain. Growth on SGal-Leu-Ura suggests a lack of dominant negative phenotype for each variant. Growth on SGal-Leu+FOA suggests in vivo activity of variant and lack of growth on SD-Leu+FOA is consistent with control of Thg1 expression on the variant vector.







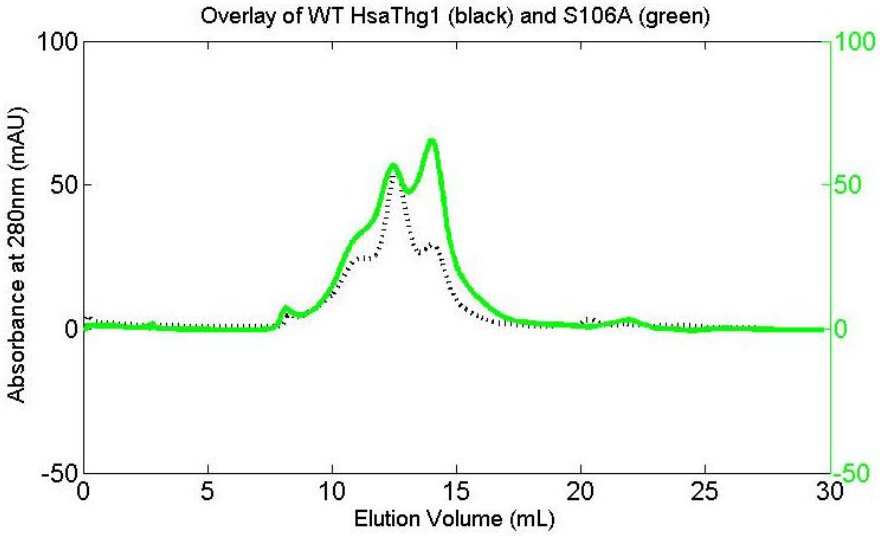


Figure 8: Chromatograms of absorbance at 280nm (AU) vs. elution volume (Ve) in mL of selected variants under 0.25 M NaCl conditions. The upper left figure displays the HsaThg1 WT chromatogram. The MW of each peak was calculated from a standard curve obtained from MW standards (BioRad) ran under the same conditions. The two major species of interest on each chromatogram are the peaks corresponding with MWs of 166 kDa and 97 kDa respectively. The 166 kDa peak is consistent with the native tetramer structure while the 97 kDa peak is an unknown lower MW species of interest. Each variant curve has the WT HsaThg1 chromatogram overlaid for direct comparison of the major peaks in each species.

Chapter 2: Identification of novel residues important for sThg1 activity

When *S. cerevisiae* is grown with a variant Thg1 that lacks the activity of the wild-type enzyme, cell viability is severely affected, and the essential nature of the Thg1 enzyme led researchers in the Phizicky laboratory at the University of Rochester to develop a screen for Thg1 alterations that lead to less dramatic growth defects, such as sensitivity to growth at higher temperature (37°C). Using this approach, a series of *thg1*^{ts} variant strains were identified in the Phizicky lab (Preston and Phizicky, unpublished data). However, an interesting observation was made that for several of these strains, in which combinations of 2 different *thg1*^{ts} alleles restored growth at high temperatures, thus eliminating the temperature sensitive phenotype as can be seen in Figure 9. Since Thg1 catalyzes a 3-step reaction for addition of G_{-1} to tRNAHis (see Figure 4), one possible explanation for this result is that each alteration affects different residues that participate separately in the G_{-1}-addition reaction, and therefore that this genetic screen might allow for the identification of novel residues that are crucial for specific steps in Thg1 catalysis. For example, it is possible that one of the variants was necessary for the nucleotide transfer step of the catalysis while the other was needed for adenylylation. A yeast strain with Thg1 monomers of only one type of variant would therefore not be able to catalyze a complete G_{-1}-addition reaction but a yeast strain that expresses both alleles could allow for a combination of each defective monomer that covers for a different step of the reaction and lead to a functional Thg1 tetramer. This approach could allow for the identification of critical residues that are not as highly conserved and may function outside of the actual active site of the enzyme. By characterizing these novel variants in vivo, non-obvious residues that could be critical for proper catalysis by SceThg1 can be identified and could lead to novel insights to proper protein function.

Creation of expression vectors through LIC cloning

To create expression vectors for each of the temperature sensitive SceThg1 variants, two approaches were followed. First primers were designed in order to create Y146H, L159H, L159F and D47Y through site-directed mutagenesis following the Quickchange protocol. Vectors containing each of these variant SceThg1 sequences were created successfully using this approach. Later, vectors containing each of the following mutations: G172D/L233S, L127S, Y8C, L63P, A111P, L222P, S148P, C134R and K59E/I70K were obtained from the collaborators in the Phizicky lab. Two primers were then designed such that the altered SceThg1 gene was amplified with an additional LIC sequence on each end of the gene that is complementary to a LIC acceptor vector, after each is treated with T4 DNA polymerase. These primers were used to amplify the SceThg1 gene on each of these variant vectors through the use of PCR following the iProof High Fidelity DNA Polymerase protocol (Biorad) and resulting in an amplified product of each variant gene of interest. LIC cloning was then used to create a pET expression vector for each of these mutants as described in the Methods section. These expression vectors were then purified from E. coli using a Qiagen miniprep kit and confirmed through the OSU Plant-Microbe Genomics Facility. The expression vectors for S148P, Y8C, L222P, C134R, K59E/I70K were created successfully through LIC cloning while the variants K96R/T99P, G172D/L233S, A111P, L63P, L127S were unable to be successfully created.

Purification of variants

The expression vectors for Y146H, L159H, L159F and D47Y sThg1 variants created in the previous step were then transformed into *E.coli* strain BL21 (DE3) pLysS (Novagen) and expressed in a 1-liter culture as described in the Methods section. These variant proteins were then purified using immobilized metal-ion chromatography and quantified using a Bradford assay as described in the Methods section. Purifications under these conditions yielded 0.5 mL of the Y146H protein at a concentration of 1.2

mg/ml and 0.5 mL of L159F protein at a concentration of 2.0 mg/ml. For both L159H and D47Y, low yields precluded an accurate determination of protein concentration. Due to the low concentration and difficulties purifying this first round of mutants, we considered the possibility that alternative conditions were needed for expression and purification of these unknown variant proteins.

Expression testing

To test the conditions needed to express stable versions of these variant sThg1 proteins, the variant Y8C and WT sThg1 were transformed into *E. coli* strain BL21 (DE3) pLysS and grown from a single colony overnight at 37°C in a 5-mL LB + amp culture. Using these overnight cultures, a 30 mL LB+ amp culture for each colony was inoculated at an OD_{600} of 0.1. This culture was grown at 37°C until the OD_{600} reached 0.4 and then each culture was split into 4X 6mL tubes such that four induction conditions could be tested for each colony: 37°C for 5 to 6 hours, 30°C for 5 to 6 hours, 18°C overnight and an uninduced culture overnight. Each of these cultures was then sonicated for two minutes at a pulse of 4 seconds with an amplitude of 80%. After repeating the sonicator treatment, the resulting cultures were then pelleted and 1 µl aliquots of crude extract and pellet for each sample were ran on an SDS Page gel. The intensity of the protein bands were compared between conditions and it was determined that the best expression was seen when the variant protein was expressed for 6 h in the 37°C induction condition (data not shown). Subsequent SceThg1 variants were grown according to this condition and yielded better purifications. Using these new induction conditions, 0.5 mL of Y8C was purified at a concentration of 5 mg/ml, 0.5 mL of S148P was purified at a concentration of 0.3 mg/mL and 0.5 mL of L222P was purified at a concentration of 0.2 mg/mL.

G_{-1} Activity assays

To measure G_{-1} activity, an assay was used in which G_{-1}-lacking $tRNA^{His}$ was 5' monophosphorylated with ^{32}P and then incubated with 0.1 mM ATP, 1mM GTP, and each purified variant Thg1 for 2 h. These reactions were then inactivated with the metal chelater EDTA and treated with RNase A, which cleaves RNA substrates 3' to pyrimidines and finally treated with calf intestine phosphatase (CIP) which cleaves any phosphate not protected by a phosphodiester bond. These reactions were then spotted on a silica-backed TLC plate and reactant products were resolved. Based on the known mechanism of Thg1 catalysis three products are expected: If the reaction does not occur then CIP will cleave the unprotected ^{32}P. If the tRNA is adenylylated but G_{-1} addition does not occur then $Ap^{32}pGpC$ is resolved from the free ^{32}P. If the reaction proceeds to completion, then the product $G^{32}pGpC$ is resolved from the other two reaction products. Each of these reaction products resolves clearly on the TLC plate and can quantified using a phosphor imager and quantification software. Using this assay, each of the variant SceThg1 proteins was tested for their G_{-1} activity on 5' ^{32}P monophosporylated $tRNA^{His}$. These results can be seen in Figure 10 and each variant shows no Gp*GpC reactant products with respect to the WT enzyme. Interestingly , each of the variants do show a small amount of product that is consistent with GTP activation of the substrate sometimes seen in these assays. This unknown product is seen in each of the reactions with the highest concentration of protein which may suggest that these variants may not be entirely deficient in the activation step of catalysis but further characterization is needed to identify the nature of the unknown product (data not shown). In order to test whether the purified variants would show activity when coupled with its complementary pair found in the yeast genetics experiments (i.e. Y8C+Y146H, Y8C+L159F, L159F+Y146H, L159F+S148P), 50 microliters of the enzyme in each pair (at matched concentrations) was incubated at 4°C overnight, then tested by the G_{-1} activity assay. Each reaction was then treated with RNase A and CIP and these products were then spotted on a silica backed TLC plate and resolved in 55:35:10 (n propanol:

NH4OH:ddH20). These results are displayed in Figure 10 and show limited Gp*GpC reaction products indicating a noticeable lack of G_{-1} addition activity as compared to the wild-type enzyme. Interestingly, both the Y8C+Y146H, and the Y8C+L159F reactions showed slight Gp*GpC reaction products in the highest concentration while each of the enzymes by themselves showed no Gp*GpC reaction products. Further, the unknown product that is consistent with GTP activation of the substrate is increased in the lower dilutions for each of the mixed reactions when compared to each of the variant enzymes by themselves but all of the SceThg1 variants show an unknown product spot in the reaction corresponding with the highest concentration of enzyme.

Discussion

Of the thirteen *thg1*^{ts} variant enzymes that were attempted to be expressed and purified, only nine were able to be cloned into the relevant expression vectors. Of those nine, all were able to be successfully expressed but only 5 purified in high enough yield to be tested for G_{-1} activity in vitro. This highlights the difficulties associated with working with these temperature-sensitive variants outside of a live cell. The majority of the highly expressed protein often appeared in the insoluble pellet before purification (data not shown). This may be due to structural instabilities present in the individual variants without their complementation pair and could be related to the observed temperature sensitivity associated with many of these variants when they are expressed by themselves in vivo.

Nonetheless, the five structurally viable variants give preliminary insight into what could be causing their defect in activity associated with each enzyme. The results in this study suggested that each of the variants, when tested alone, show some capability for potential activation of the G_{-1} lacking $tRNA^{His}$ at the highest amount of enzyme (0.3 ug – 5 ug). Perhaps even more interesting, when the variants are mixed with their complementary pair, this unknown product is present in multiple reactions, even with lower concentrations of the purified enzymes. This may indicate a trapped

intermediate in a single step of enzyme catalysis that could be crucial for understanding structural transition states in enzyme catalysis. Furthermore, reactions performed in the presence of the Y8C + Y146H and Y8C+ L159F Thg1 variants clearly both show some capacity for G_{-1} activity, while each variant when tested individually does not. Clearly activity between complementary pairs is able to be recapitulated in vitro and may indicate a critical role for the residue Y8 in the activity of the enzyme that had not been appreciated in any previous experiment, due to the overall lack of conservation of the Y8 residue in Thg1 family enzymes.

Viewing the crystal structure of hThg1, which shares significant homology with the SceThg1, the *H. sapiens* counterparts to the residues explored in this study are each located close to the known catalytic residues in the active site of the enzyme (Figure 11). Of particular note is residue Y8 which is located in a highly dynamic region of the enzyme. The relative proximity of this residue to the active site and freedom of movement could allow this residue to play a critical role in substrate recognition or binding. Interestingly, Y146 (Y143 in *H. sapiens*), L159 (L156 in *H. sapiens*) and Y8 are all located in close proximity to each other and the active site but both Y146 and L159 are on a single monomer, while Y8 occupies a second monomer in the final multimeric formation of the enzyme. If a chimeric SceThg1 comes together in solution with a monomer of either Y146H or L159F mixed with a monomer of Y8C, a novel conformation could be formed to recapitulate native activity. These data have helped support the importance of these residues in the final functional enzyme. Further characterization of the Y8C variant and other mutations of this region of the enzyme may better indicate its role in the overall activity of the enzyme and could lead to novel insights of the molecular determinants of catalysis by Thg1 enzymes.

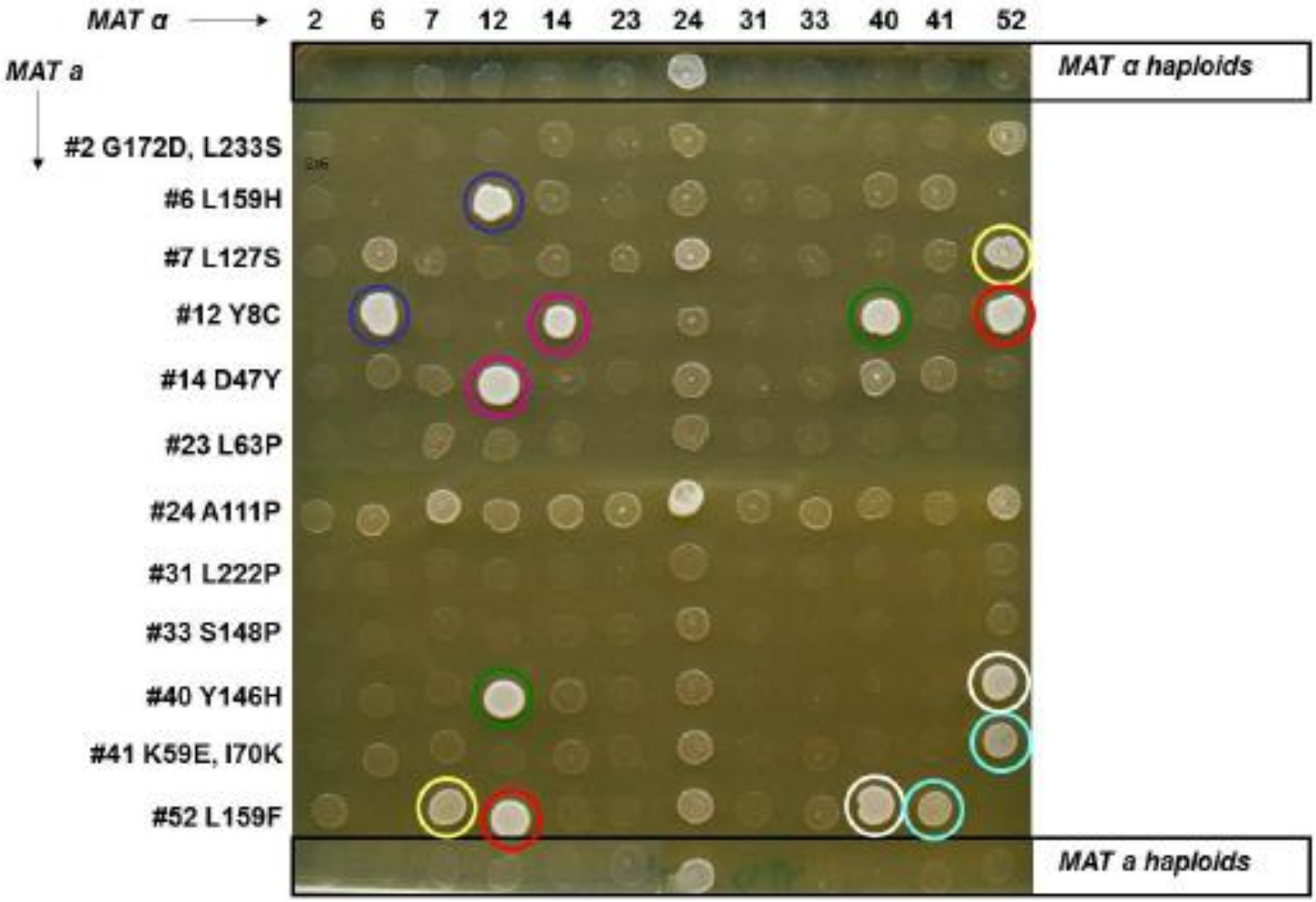


Figure 9: Yeast genetics assay showing complementation of temperature sensitive sThg1 variants. The numbers for each variant along the side correspond the numbers along the top and cross-sections represent co-expression of each corresponding variant Thg1. Experiment was conducted by collaborators in the Phizicky lab. (Preston & Phizicky unpublished).

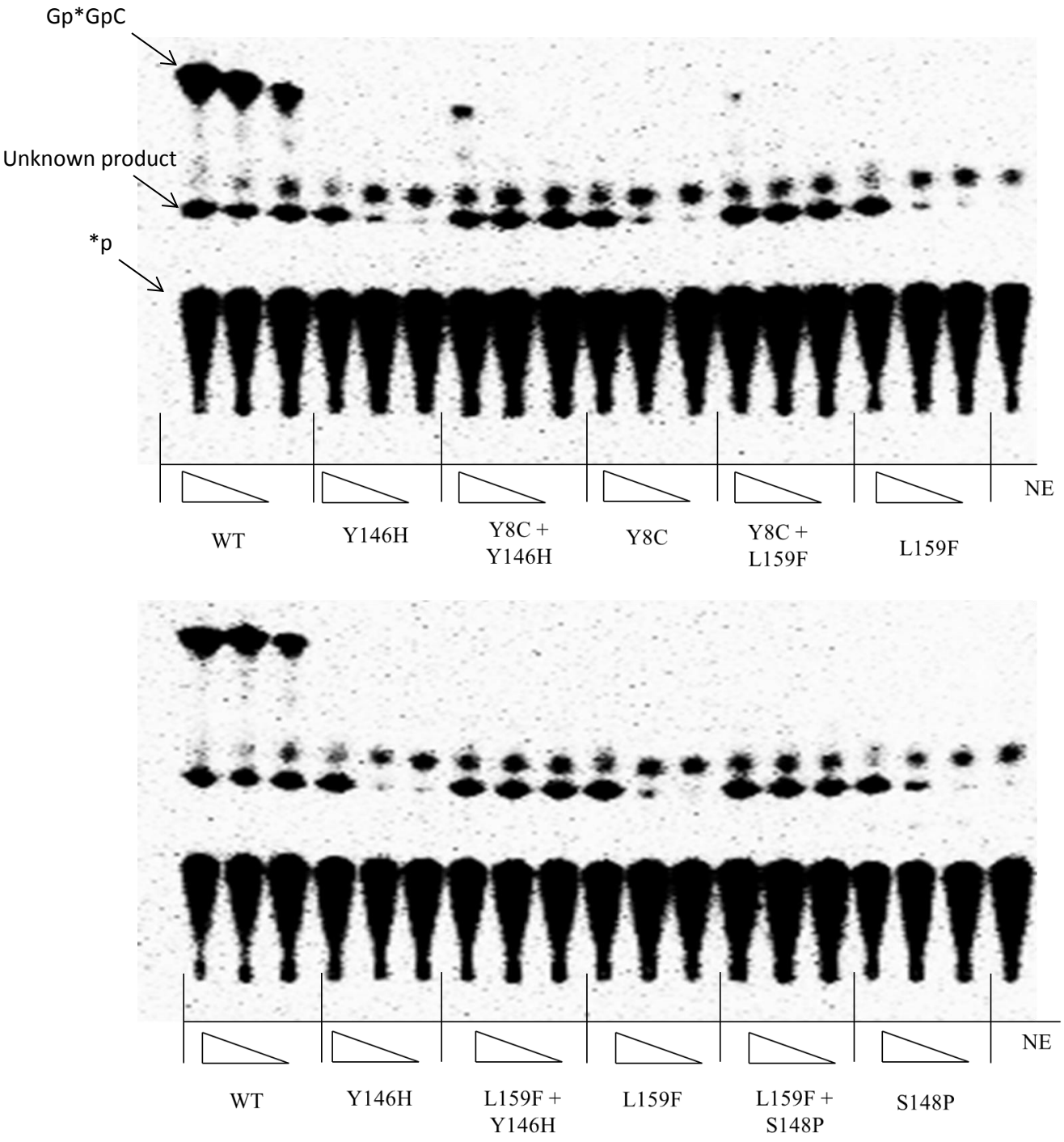


Figure 10: G-1 addition activity assay. Each enzyme was mixed with P^{32} labeled $tRNA^{His}$ in 0.1 mM ATP and 1 mM GTP for 2 hours, digested with RNase A and CIP, and run on a silica TLC plate in 55:35:10 n-propanol:NH_4OH:ddH_2O as the solvent. The NE control indicates the location of free phosphate label while the spot migrating directly below indicates an unknown reaction product consistent with GTP activation while the higher migrating spot indicates a Gp*GpC reaction product. Each enzyme was tested in 5-fold serial dilutions as indicated above.

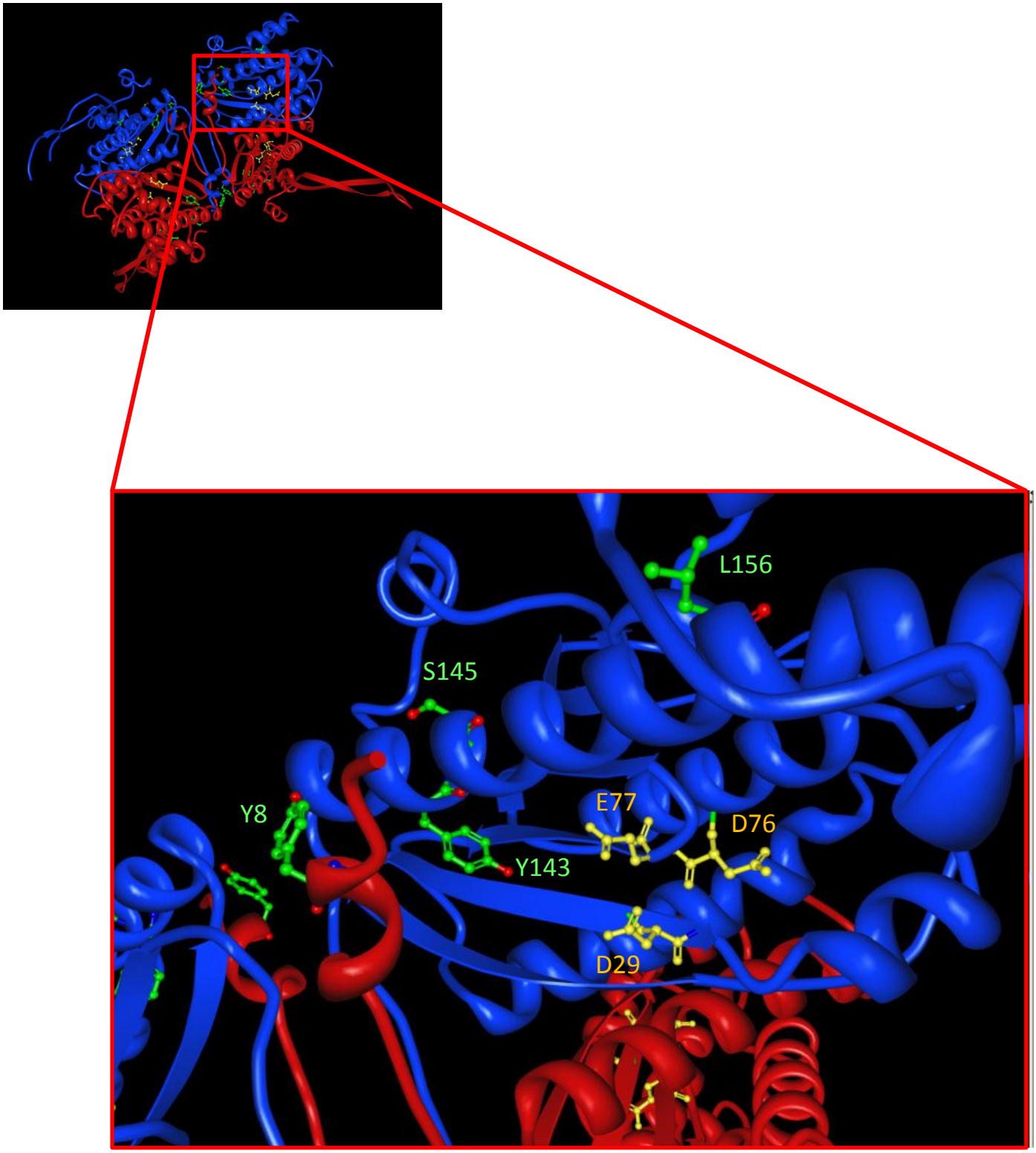


Figure 11: hThg1 crystal structure highlighting the relative proximity of the residues investigated in this project (green) to the known catalytic residues of the WT enzyme (gold). Sequence alignment between *H. sapien* and *S. cerevisaie* Thg1 allow for identification of counterpart residues between the two enzymes. In this case the human residue Y143 is equivalent to yeast Y146, human S145 is equivalent to yeast S148 and human L156 is equivalent to yeast L159. Asymmetric monomers are indicated in blue and red.

Chapter 3: Residues required for differential HisRS specificity between two eukaryotic organisms

Histidyl-tRNA synthetase (HisRS) is an essential enzyme needed to aminoacylate $tRNA^{His}$ for protein translation. In species from all three domains of life, an additional guanine residue at the -1 position (G_{-1}) of $tRNA^{His}$ is used as an essential identity element for substrate recognition and consequent aminoacylation by its cognate HisRS. Investigations of diverse eukaryotes, such as *Trypanosoma brucei* and *Acanthamoeba castellanii* revealed that the nuclear-encoded $tRNA^{His}$ in these species lacks a G_{-1} residue, and the native HisRS is capable of efficient aminoacylation independent of this critical identity element (11). Biochemical and genetic studies implicated a region of HisRS known as Motif II in G_{-1} recognition, although the precise molecular details of this interaction remain to be fully determined. Sequence alignments indicate differences between the motif II sequences from non-canonical HisRS enzymes from *A. castellanii* and *T. brucei* and their G_{-1}-dependent counterparts, suggesting that this region may be important for the alternative biochemical properties exhibited by these enzymes (Figure 12).To investigate whether these residues play a role in alternative substrate recognition, a 13 residue region in motif II of the HisRS from each species was exchanged between the two organisms. These variant proteins were expressed, purified, and tested with in vitro aminocylation assays on both G_{-1}-containing and G_{-1}-lacking substrates. Additionally, a smaller 4 residue region and an individual residue thought to be potentially active in G_{-1} recognition were also exchanged between the two organisms and similarly analyzed. Taken together, this investigation will help to identify how two different HisRS enzymes are capable of recognizing a single correct substrate among the other structurally similar tRNA species present in each organism. This information could also lead to the development of selective inhibitors that could take advantage of the differences between the pathogen *T. brucei* and *H. sapiens* HisRS activities.

Creation of expression vectors

Variants were selected by sequence alignment between the canonical and non-canonical HisRS enzymes. Figure 12 displays this alignment between the organisms. Representative organisms were chosen for each type, *S. cerevisiae* for canonical HisRS and *A. castellani* for non-canoncial HisRS. For the first round of variants, primers were designed to exchange Motif II (indicated by a red square in Figure 12) between the two sequences such that a *S. cerevisiae* HisRS was created with the *A. castellani* Motif II (AcinSc) and vice versa (ScinAc). Using parent expression vectors for each WT enzyme, the variant sequences were created through Phusion Site-Directed Mutagenesis as described in the methods section. For the second round of variants, residues were carefully chosen in a region of Motif II that is thought to be an interacting loop with the G_{-1}, or lack thereof, in the tRNA bound to the HisRS and is indicated by a green square in Figure 12. The first of these residues chosen were the *S. cerevisiae* residues R150 and D151 (indicated by the orange arrow in Figure 12) due to their highly conserved nature in the canonical HisRS organisms. The corresponding residues in *A. castellani,* F168 and E169, and *T. brucei,* YE, are each large amino acids and the similarities among each type of HisRS suggest these residues to be good candidates responsible for the difference in specificity between the two types of enzymes. The variants AcFEtoRD and ScRDtoFE were attempted to be created based on these residues with Phusion Site-Directed Mutagenesis but were unsuccessful. The second set of residues chosen in this second round was the QPAM region in canonical HisRSs and corresponding TV region in non-canoncial HisRSs (indicated by the purple box in Figure 12). Once again, the conserved nature of these residues among each type of HisRS and conserved difference between each type of HisRS makes this region another excellent candidate. The variant AcTVtoQPAM was successfully created using these residues as guides but the variant ScQPAMtoTV was not successfully created. Finally, the *S. cerevisiae* residue M160 (as indicated by the blue arrow in Figure 12) was chosen due to its highly conserved nature of the non-polar amino acids found at this location in canonical HisRS species. The corresponding residue in non-

canonical HisRS species is a conserved arginine residue and the conserved nature of this residue among each HisRS type but clear difference between each HisRS type made this residue another fantastic candidate. Both the AcRtoM and ScMtoR variants were created successfully using these residues as guides. Successful sequences were confirmed by Genewiz.

Purification of HisRS variants

To purify each protein of interest, variant plasmids were transformed into XL1-Blue competent *E. coli* cells and grown on LB+Amp plates. Single colonies were chosen for each variant and a 5 ml culture of LB+Amp was grown overnight. The DNA was then purified using a Miniprep kit. The variant DNA was then transformed into BL21 cells and a 1-liter LB+Amp culture was grown at 37°C until the OD_{600} read 0.4. The culture was then induced with 1mM IPTG and grown for 6 hours at 37°C. The cells were then harvested and the protein purified from a 0.5 L pellet using metal-ion chromatography as described in the methods section. Variant proteins were confirmed by SDS-PAGE gel and concentration was determined through the use of a BioRad Bradford assay. Purification under these conditions yielded 0.5 ml of 5.0 mg/ml for AcinSc, 0.5 ml of 20.0 mg/ml for ScinAc, 0.5 ml of 1.0 mg/ml for ScMtoR, 0.5 ml of 16.0 mg/ml for AcRtoM, and 0.5 ml of 10.0 mg/ml for AcTVtoQPAM.

HisRS activity assays

The sequences for both G_{-1}-containing (Gu2) and G_{-1}-lacking (Gu3) $tRNA^{His}$ were obtained from the lab stock and transcribed as described in the methods sections. The transcribed tRNA was then 3' labeled according to the conditions described in the methods section. Each variant enzyme was then incubated in a separate reaction with each labled substrate (Gu2CCp*A and Gu3CCp*A), histidine and inorganic pyrophosphatase which serves to shift the equilibrium of the reaction towards aminoacylation as described in the methods section. Each variant was diluted by serial dilution in two five-fold dilutions in

HisRS dilution buffer. This resulted in 3 reactions (undiluted, 1:5, 1:25) for each variant and substrate combination. The results can be seen in Figure 13. The lower left panel in Figure 13, displays both the ScinAc variant and the AcinSc variant as compared to the Ac WT HisRS on Gu3CCp*A substrate. As can be clearly seen, both variants show no visible activity on this substrate as compared to WT. The upper left panel in Figure 13 displays ScinAc's, AcinSc's, AcRtoM's, and ScMtoR's activity on Gu2CCp*A as compared to WT Ac HisRS. Both AcRtoM and ScMtoR display WT level activity on this substrate while both AcinSc and ScinAc show no discernible aminoacylation activity. Finally, the upper right panel in Figure 13 displays AcaRtoM's, ScMtoR's, and AcTVtoQPAM's activity on Gu3CCp*A as compared to WT Ac HisRS, as well as AcTVtoQPAM's activity on Gu2CCp*A. Both AcRtoM and ScMtoR show comparable levels of aminoacylation activity as compared to the WT. AcTVtoQPAM shows substantially reduced activity on Gu3CCp*A and no discernible activity on Gu2CCp*A.

Discussion

It was hypothesized that, by switching the Motif II of a non-canonical HisRS with that of a canonical HisRS, the specificity of each enzyme could be switched. In this case the variant AcinSc would lose its preference for G_{-1} containing $tRNA^{His}$ while ScinAc would gain a preference for G_{-1} containing $tRNA^{His}$. Although each variant purified in relatively high yields (5.0 mg/ml and 20.0 mg/ml respectively) suggesting structural stable protein, neither protein showed any discernible activity on either substrate. This is likely due to the large size of the region chosen for the swap. It is likely that too much of the structure was changed from the WT enzyme for each representative organism resulting in inactive enzymes. In order to combat this, the second round of more selected variants was created. These variants allowed for much smaller swaps to be conducted based on the rationale that residues whose properties were conserved among each type of HisRS but differed between them were likely candidates to be involved in recognition of G_{-1} on the substrate tRNA. The *A. castellani* variants with *S. cerevisiae*

residues showed little change in specificity for substrate. The AcRtoM variant showed clear activity on either substrate and mirrored it's WT counterpart, whereas the AcTVtoQPAM displayed a reduced capacity for activity showing no activity on G_{-1}-containing substrate and limited activity on G_{-1}-containing substrate. These results suggest that the changing the selected residues is not enough to cause the *A. castellani* enzyme to selectively choose G_{-1}-containing substrate and further exploration of the interacting loop may give better insights. However, the *S. cerevisiae* variants with *A. castellani* residues tell a different story. ScMtoR shows activity on both G_{-1}-containing and G_{-1} lacking tRNA substrates. As a canonical HisRS, the *S. cerevisiae* will only aminoacylate G_{-1}-containing substrate but by changing a single conserved residue from the neutral amino acid methionine to the basic amino acid arginine this specificity for G_{-1} containing substrate is eliminated. These results suggest that the residue M160 is likely directly involved in susbstrate recognition by canonical HisRS's and represents the first example of a molecular basis for this unique activity. Taken together with the AcRtoM activity, these results suggest that this residue is not the only residue critical for this specificity and further studies will be needed to better map out the site of recognition for both canonical and non-canonical HisRS's. By mapping and understanding the molecular basis for this differential specificity between canonical and non-canonical HisRS's, selective inhibitors of the non-canonical HisRS can be developed. *T. brucei* is the organism that causes African Sleeping Sickness, a lethal disease that is prevalent in sub-Saharan Africa. Current treatments are either not widely available or not very effective and the development of a new drug that acts on this fundamental difference between the canonical *H. sapien* HisRS and the *T. brucei* HisRS could save thousands of lives every year.

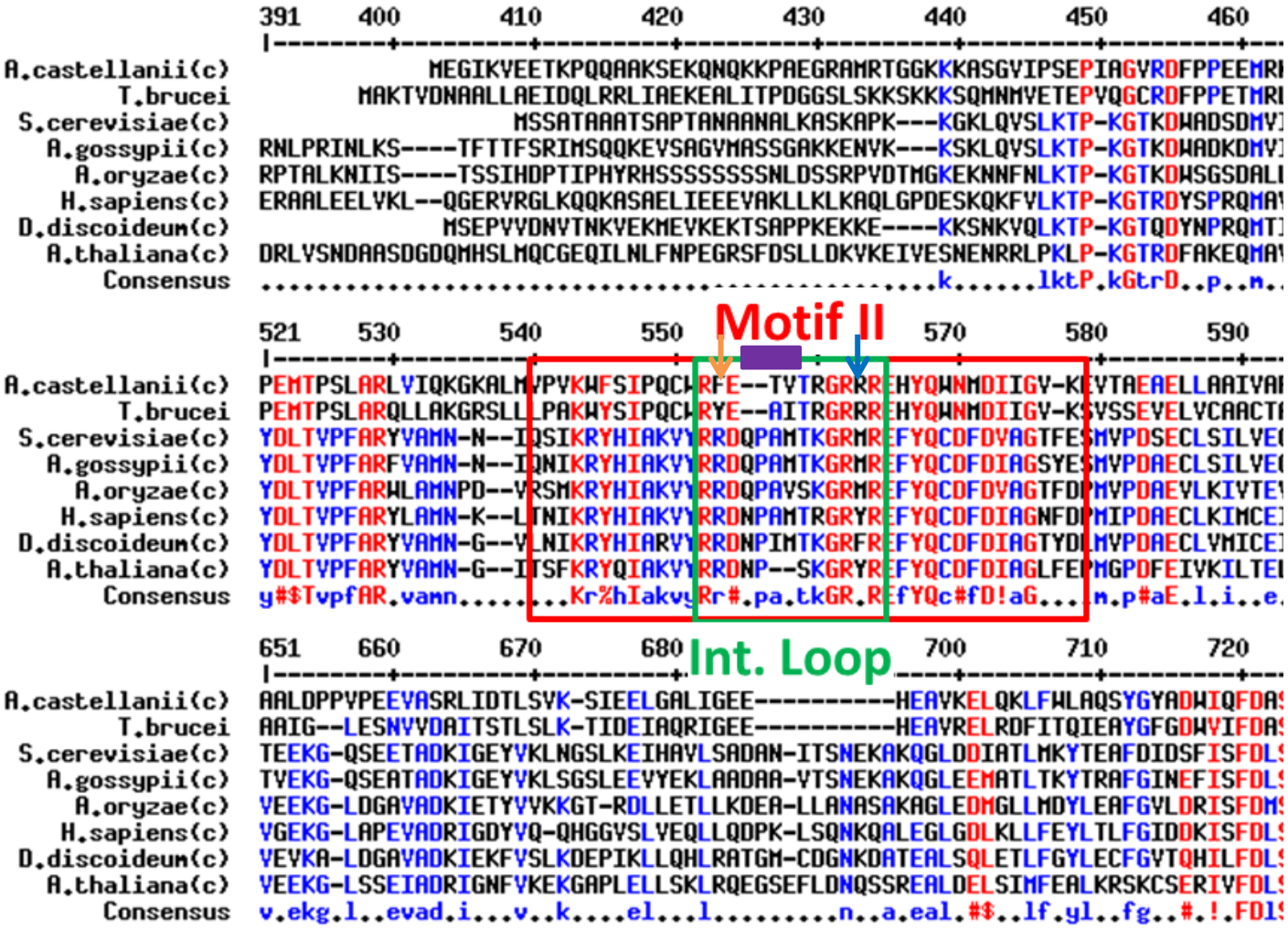
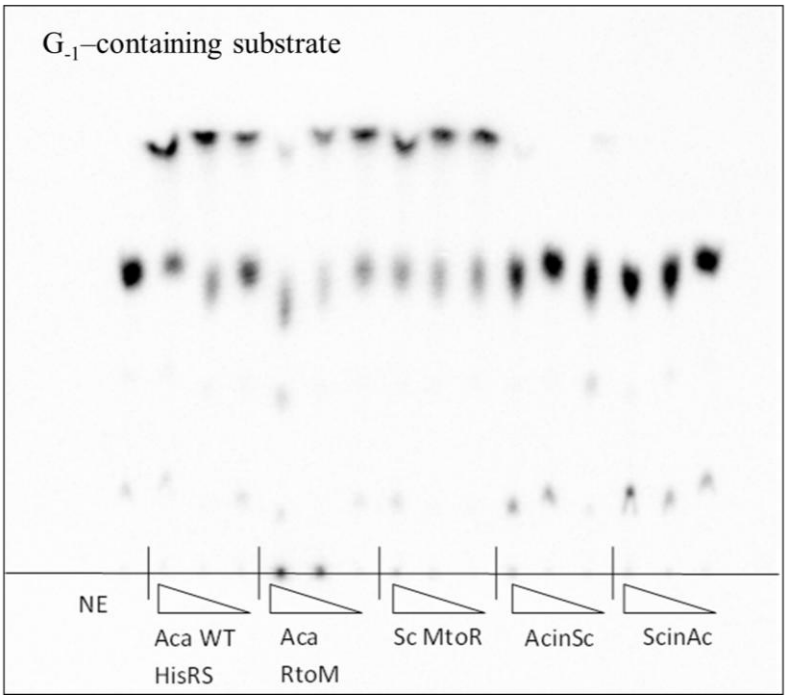
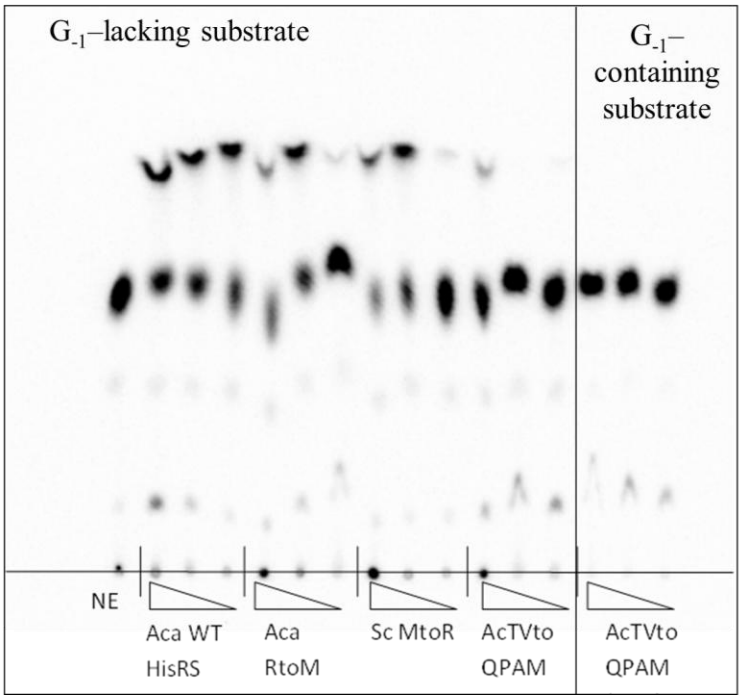
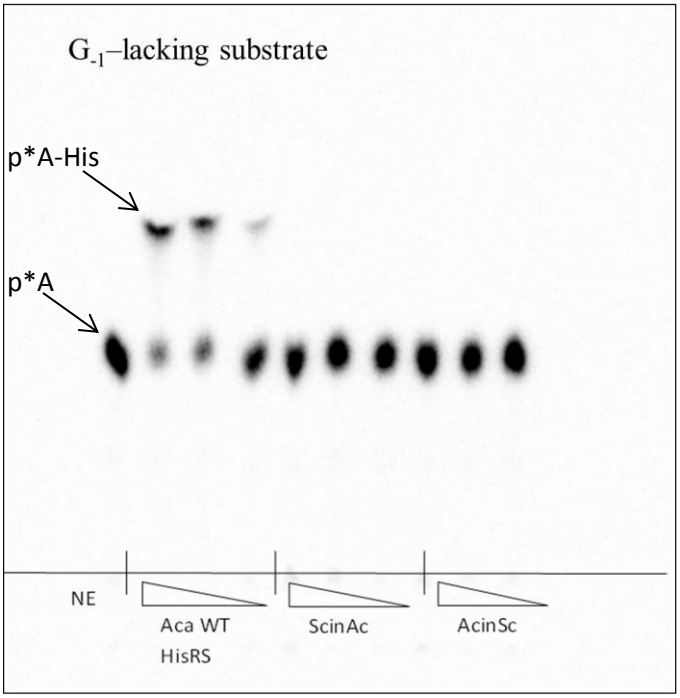


Figure 12: Sequence alignment of both canoncial and non-canoncial HisRS enzymes. The red box indicates Motif II which is hypothesized to be critical for G_{-1} recognition. The green box indicates the specific region within Motif II that is thought to directly interact with the G_{-1}.The orange arrow indicates *S. cerevisiae* residues R150 and D151 and corresponding residues F168 and E169 in *A. castellani.* The purple box indicates the TV region in in *A. castellani* and corresponding QPAM region in *S. cerevisiae.* The blue arrow indicates the *S. cerevisiae* residue M160 and corresponding R in non-canonical HisRS species. Variant chimeric HisRS's were made by swapping these residues between each organism type.







***A. castellani* Variants [conc.]**	Description
ScinAc [20.0 mg/ml]	Swap out with *S. cerevisiae* Motif II
Ac TVtoQPAM [10.0 mg/ml]	Small swap out with *S. cerevisiae* interacting loop
Ac RtoM (R176M) [16.0 mg/ml]	Replacement of *A. castellani* residue with *S. cerevisiae* counterpart
***S. cerevisiae* Variants**	Description
AcinSc [5.0 mg/ml]	Swap out with *A. castellani* Motif II
Sc MtoR (M160R) [1.0 mg/ml]	Replacement of *A. castellani* residue with *S. cerevisiae* counterpart

Figure 13: Aminoacylation assays for each variant on both G_{-1}–containing and G_{-1}–lacking $tRNA^{His}$. Each reaction was incubated with CCp*A labeled tRNA and subjected to the reaction conditions described above. Resulting products from the P1 digest were spotted on a PEI cellulose plate and ran in 5(glacial acetic acid):10(1M NH_4Cl):85 (ddH_2O). The plates were then imaged through a Phosphor Imager. The higher migrating products are p*A-His consistent with aminoacylation. The lower migrating products are p*A and suggest a lack of aminoacylation. Each plate was run with a no enzyme (NE) control lane and a WT control (Aca WT HisRS). G_{-1}.The table displays the variants successfully created from Phusion site-directed mutagenesis on corresponding WT vectors and their respective concentrations. Each variant was diluted from stock concentration in 5-fold serial dilutions for titration of activity

Materials and Methods:

Site Directed Mutagenesis

Thg1 variants were created from either the JEJ111 yeast expression vector containing WT HsaThg1 gene with a LEU2 marker, or the JEJ109 pET expression vector containing the WT SceThg1 gene. 10 ng of the Thg1 DNA and 2 oligonucleotides (125ng each) designed to create the variant of choice were used with the Quick Change Site-Directed Mutagenesis protocol (Agilent Technologies) with a minor change. Instead of adding both sense and anti-sense oligos to the parent DNA in a 50 ul reaction, two separate 25 ul reactions were completed for each variant, one for the antisense oligo and one for the sense oligo. Each reaction was then denatured for 1 minute at 95°C followed by 10 cycles consisting of 1 minute of denaturation at 95°C, 1 minute and 30 seconds of annealing at 55°C, and 1 min per Kb of the vector of elongation at 68°C. Each matching sense and antisense oligo reaction was then mixed for 18 more identical cycles in a Bio-Rad Thermal Cycler. The resulting product was then treated with Dpn1 restriction enzyme to remove any remaining parent vector and transformed into XL1-blue competent *E. coli* cells. A single colony from this transformation was picked and grown in a 5 mL LB+Amp culture. Finally the resulting variant vector was purified with a Miniprep kit (Qiagen) and confirmed by sequence at the OSU Plant-Microbe Genomics Facility.

HisRS variants were created using Phusion Site-Directed Mutagenesis using 50 ng of a parent vector containing either WT *A. castellani* or WT *S. cerevisiae* HisRS gene in a PET expression system. Oligos were designed for each of the variants and 100mM of each oligo was phosphorylated using T4 polynucleotide kinase(NEB) with the appropriate NEB buffer and 10mM ATP in a 30 ul reaction. These phosphorylated oligos where then diluted to 10mM and used with the appropriate vector in the reaction conditions outlined by the Phusion Site Directed Mutagenesis protocol. The cycling conditions used for the PCR were as follows: 1 minute denaturation at 98°C followed by 5 cycles of 1 minute

denaturation at 98°C, 30 seconds annealing at 40°C, 4 minutes elongation at 72°C. This was then followed by 25 cycles consisting of 10 seconds of denaturation at 98°C, 30 seconds of annealing at 60°C, 3 minutes of elongation at 72°C. Finally a 10 minute final elongation was conducted at 72°C. The resulting PCR product was confirmed on an analytical 1% agarose gel. Once confirmed, a 8 ul of each PCR reaction was ligated using 10 U/ul T4 DNA Ligase (USB) following the reaction conditions recommended by the manufacturer with the exception of incubation at 37°C for 2 hours rather than the recommended 18°C overnight. The entire ligation products were then incubated with 1ul of Dpn1 at 37°C for 2 hours to remove any background WT vector. Finally 2 ul of the Dpn1 reaction was transformed into XL1-blue competent *E. coli* cells. A single colony from this transformation was picked and grown in a 5 mL LB+Amp culture. Finally the resulting variant vector was purified with a Miniprep kit (Qiagen) and confirmed by sequence from Genewiz.

LIC Cloning

First, the gene inserts were amplified from each vector of interest by following the IProof DNA Polymerase (BioRad) protocol with the appropriate primers. The resulting amplified genes were confirmed by size and quantified through an analytical 1% agarose gel using a 1 Kb plus ladder (BioRad) to determine concentration. The PCR products were then treated with T4 DNA polymerase according to the supplier's reaction conditions but only in the presence of 25 mM dTTP. The AVA421 vector (pET expression vector with complementary LIC sequence) was first treated with NruI and then Pme1 according the reaction conditions specified by the supplier. The cut AVA421 vector was then purified through gel purification using the Qiaquick Gel Purification Kit and following the provided protocol (Qiagen). The cut AVA421 vector was then treated with T4 DNA polymerase in reaction conditions determined by the supplier of the enzyme but only in the presence of 25 mM dATP. A minimum of 0.005 picomoles of T4-treated insert was then incubated with a minimum of 0.0025 picomoles of T4 treated

insert for 5 minutes at RT. EDTA is then added to the reaction for a final concentration of 6.25 mM and incubated for additional 5 min. The final reaction was then transformed into XL1-Blue cells and grown on LB+Amp plates. Finally a single colony from each transformation were picked and grown in a 5 mL LB+Amp culture and the resulting variant vectors were purified with a Miniprep kit (Qiagen) and confirmed by sequence from the OSU Plant-Microbe Genomics Facility.

Purification of recombinant protein

Once confirmed by sequence, the variant DNA plasmids were transformed into BL21(DE3) pLysS (Novagen). A single colony for each variant-expressing strain was grown in 5 ml of liquid LB plus 100 µg/ml ampicillin (amp) for 6 hours. This culture was diluted 1:500 in a 25 ml LB+amp culture and grown overnight at 37°C with shaking at 250 rpm (all cultures were grown at 37°C with shaking at 250 rpm unless otherwise indicated). After overnight growth, the OD600 of the 25 ml culture was measured and was then used to inoculate to OD_{600}= 0.01 in a 1 liter culture of LB + amp. The 1 liter liquid culture was then allowed to grow until the OD_{600} reached 0.4 consistent with the log phase of growth in *E. coli*. At this point the BL21 cells were induced to begin expressing the variant Thg1 by adding 1mM IPTG from a 100 mM fresh solution. The temperature was then changed to 18°C and the culture was allowed to grow up to 20 hours. The cells from the 1 L culture were harvested and half of the culture was stored at -80°C. The other half was used in purification. The His6-tagged protein was purified using immobilized metal ion affinity chromatography. All of the following steps were performed at 4°C, unless otherwise indicated. The cell pellet was resuspended in 20 mL of Buffer A (20mM HEPES 7.5, 4mM MgCl2, 10% glycerol, 0.0005 M beta-mercaptoethanol (BME)) + 1M NaCl and lysed by French Press. After centrifugation to remove cellular debris, an equal volume of Buffer A was added to the soluble crude extract to yield a final concentration of 0.5 M NaCl. The crude extract was added to 2mL of TALON resin (Clontech), pre-equilibrated in Buffer A + 0.5 M NaCl, and allowed to mix gently for 20 min. The resin

was repeatedly washed by centrifuging for 2 min at 2000 rpm, and removing wash buffer, followed by addition of each new buffer and gentle mixing. A total of 3 washes (each with greater than 10x the resin volume) were completed with Buffer A + 0.5 M NaCl and then 3 washes with Buffer A + 0.5 M NaCl and 10mM imidazole pH 7.7. At this point the slurry was poured into a disposable column, and the protein was eluted using 10ml of Buffer A + 0.5 M NaCl and 250mM imidazole. The protein was eluted into 10 separate 1 ml fractions in 2 ml microtubes. These fractions were then tested for the presence of protein using BioRad Protein assay. The fractions with the most protein in them were pooled and dialyzed in either a Thg1 50% glycerol dialysis buffer (20mM tris 7.5, 0.5M NaCl, 1mM MgCl2, 50% glycerol, 500mM EDTA pH 8.0, 1M DTT) or a HisRS 50% glycerol dialysis buffer (20mM tris 7.5, 0.05M NaCl, 4mM MgCl2, 50% glycerol, 500mM EDTA pH 8.0, 1M DTT) overnight and stored at -20°C. The next day the protein was quantified by using a BioRad Protein assay with IgG standards and was confirmed by running an SDS-PAGE gel (BioRad).

Transcription of tRNA^{His}

A plasmid (Gu3) containing the coding sequence for tRNA^{His} under control of the T7 RNA polymerase promoter was linearized with NsiI, and used for run-off transcription with T7 RNA polymerase. The digested DNA was purified by phenol:chloroform extraction followed by ethanol precipitation, and resuspended in 50 µl ddH2O. The resulting DNA was quantified using 260 nm absorbance, where one absorbance unit was equal to 50 µg/µl concentration. The digested DNA was treated with T4 DNA polymerase (according to the manufacturer's instructions) to remove 3' overhanging nucleotides that inhibit in vitro transcription. The DNA was then purified and quantitated again as above. The entire digest was used for in vitro transcription reactions containing 0.1 mg/ml digested DNA, and T7 RNA polymerase, under standard conditions (*10*). The RNA was purified by

denaturing polyacrylamide gel electrophoresis and resuspended in TE 7.5 for subsequent labeling reactions, as appropriate for the enzymatic assays to be performed with each variant enzyme.

5' Labeling of $tRNA^{His}$ for G_{-1} activity assay

$tRNA^{His}$ transcript was treated with 10 U/ul CIP (NEB) following the reaction conditions recommended by the supplier in order to remove any 5' phosphates. The resulting $tRNA^{His}$ was then purified through phenol: chloroform extraction and ethanol precipitation. In order to add a labeled phosphate to the 5' end, the tRNA was treated with T4 PNK in the presence of $\gamma[^{32}P]$-ATP at 37 °C for 30 minutes. Following heat inactivation at 72 °C for 10 minutes, the substrate was centrifuged in a BioGel P6 tube, removing excess ATP. The final substrate was then gel-purified in a 10% polyacrylamide gel in the presence of 4 M Urea.

G_{-1} addition activity assay

Each of the variant enzymes to be tested were titrated in either 5 or 10-fold dilutions in 20mM Tris pH 7.5, 0.5 mg/ml BSA, 500 mM NaCl and 1 mM DTT dilution buffer. These dilutions were then incubated with the 5' mono-^{32}P labeled $tRNA^{His}$ in 25mM Hepes pH7.5, 10mM $MgCl_2$, 3 mM DTT, 125 mM NaCl, 0.2 mg/ml BSA, 0.1 mM ATP, 1mM GTP for 2 hours at RT. These reactions were then inactivated with EDTA and treated with 5 ug of RNase A for 10 minutes at 50 °C. These reactions are then treated with 0.5 Units of CIP (Roche) removing any unprotected phosphates. These reactions where then spotted on a silica-backed TLC plate and reaction products were resolved in 55:35:10 (n-propanol: NH_4OH, ddH_2O), visualized by a phosphor imager and quantified using ImageQuant software.

Size-exclusion Chromatography

The previously purified WT and variant hThg1 proteins were filtered using 1.5 ml Eppendorf spin tubes and 1 L of 50 mM $NaHPO_4$ + 250 mM NaCl pH 7.2 was created from a mixture stock solutions of monobasic and dibasic sodium phosphate. A Superdex 200 10/300 GL column was used with an AKTA FPLC (GE healthsciences) and 500 ul of each protein was loaded (0.712 mg – 8 mg) and ran in the sodium phosphate buffer. The absorbance at 280 nm was read as the proteins eluted out of the column and the elution volumes were determined. These volumes were compared with a BioRad molecular weight standard to determine relative molecular weight of the elution fractions.

Yeast Complementation Assay

To test the function of Thg1 variants in yeast, a plasmid containing Thg1 (yeast or human) expressed under control of a galactose inducible promoter with a *LEU2* marker was used. This plasmid was used as the template for creating a series of site-directed variants by Quick Change (Agilent) to be tested for their ability to support growth of the yeast thg1Δ strain. The correctly sequenced Thg1 variant plasmids [*CEN PGAL-thg1* LEU2*] were transformed into yeast strain JJY20 (relevant genotype: *mat a, thg1Δ::kanR,leu2Δ,ura3Δ[CEN PTHG1-THG1 URA3]*). This strain contains a covering plasmid with the wild-type *THG1* under control of the PTHG1 promoter. After transformation of the *LEU2* variant plasmids and selection on a synthetic dextrose (2%) media lacking leucine and uracil (SD-leu-ura) , single colonies were taken from this plate and streaked onto a 5-fluoroorotic acid (FOA)+ 2% Galactose plate to ensure that only strains with the second *LEU2* vector grow. The galactose promotes the production of the mutant *thg1* gene and the FOA makes the expression of the *URA3* gene toxic which produces selective pressure for the cell to lose the *URA3*-containing wild-type *THG1* covering plasmid. Such an approach ensures that the Thg1 variant gene on the LEU2 plasmid is the only source of Thg1 in the cell,

and growth on SGal-leu+FOA indicates that the variant Thg1 is functional, if no growth is observed, then the variant Thg1 is not functional.

3′ Labeling of $tRNA^{His}$ for amino acylation assay

The 3' end of $tRNA^{His}$ was labeled according to the methods presented by Ledouz and Uhlenbeck (12). To accomplish this goal, 1 uM of cold $tRNA^{His}$ transcript was incubated with CCA-adding enzyme in 25 microliter reaction in buffer (glycine pH 9 50 mM, $MgCl_2$ 10mM,alpha-ATP, sodium pyrophosphate 50 uM) for 15 minutes at 37°C in order to favor the 3'-5' exonuclease activity of the enzyme. CTP was then added to the reaction for a total final concentration of 2 uM along with 500 units of inorganic pyrophosphatase. The now labeled tRNA was then purified through phenol:chloroform extraction and ethanol precipitation resulting in a final $tRNA^{His}$ labeled on the CCA end in the following way: CpCp*A.

Aminoacylation assay

The following protocol was adapted from Francklyn et al. (13). Prior to incubation with histidine and the variant HisRS of interest, the CpCp*A labeled $tRNA^{His}$ was folded by heating to 75°C in a heat block for 5 minutes in tRNA buffer (50mM Hepes pH 7.5, 20 mM KCl). Once heated, $MgCl_2$ was added to the buffer to a final concentration of 10 mM and the heat block was removed from the heating element, allowing the tRNA to slowly cool to room temperature for 30 minutes. The folded tRNA was then added to buffer containing histidine (50mM Hepes pH 7.5, 10mM $MgCl_2$, 20 mM KCl, 8 mM DTT, 5 mM ATP, 0.08 mM L-Histidine, 20 U/ml inorganic phosphatase) for a total reaction volume of 19 ul. Each variant HisRS (1 mg/ml – 16 mg/ml) was then diluted in 2-4 fivefold serial dilutions in HisRS dilution buffer (20 mM Tris pH 7.5, 0.5 mg/ml BSA, 5 mM NaCl, 0.2 mM DTT). To initiate the reaction, 1 ul of each variant enzyme was added to the amino acid/ tRNA mixture and incubated at 37°C for 15 minutes. Afterwards, a 2 ul aliquot of the reaction was treated with 2 ug of P1 Nuclease in digestion buffer (20 mM NaOC, .2 mM

$ZnCl_2$) for a total reaction volume of 6 ul. The reaction products were spotted on a PEI cellulose TLC plate and run in 5:10:85 (glacial acetic acid: 1m NH_4Cl: ddH_2O). The plate was then imaged using a phosphor imager and quantified using ImageQuant software.

References

1. Giege, R., Sissler, M., and Florentz, C. (1998) Universal rules and idiosyncratic features in tRNA identity, *Nucleic Acids Res 26*, 5017-5035.

2. Schimmel, P., Giege, R., Moras, D., and Yokoyama, S. (1993) An operational RNA code for amino acids and possible relationship to genetic code, *Proc Natl Acad Sci U S A 90*, 8763-8768.

3. Cooley, L., Appel, B., and Soll, D. (1982) Post-transcriptional nucleotide addition is responsible for the formation of the 5' terminus of histidine tRNA, *Proc Natl Acad Sci U S A 79*, 6475-6479.

4. Gu, W., Hurto, R. L., Hopper, A. K., Grayhack, E. J., and Phizicky, E. M. (2005) Depletion of Saccharomyces cerevisiae tRNA(His) guanylyltransferase Thg1p leads to uncharged tRNAHis with additional m(5)C, *Mol Cell Biol 25*, 8191-8201.

5. Gu, W., Jackman, J. E., Lohan, A. J., Gray, M. W., and Phizicky, E. M. (2003) tRNAHis maturation: an essential yeast protein catalyzes addition of a guanine nucleotide to the 5' end of tRNAHis, *Genes & Development 17*, 2889-2901.

6. Hyde, S. J., Eckenroth, B. E., Smith, B. A., Eberley, W. A., Heintz, N. H., Jackman, J. E., and Doublie, S. (2010) From the Cover: tRNAHis guanylyltransferase (THG1), a unique 3'-5' nucleotidyl transferase, shares unexpected structural homology with canonical 5'-3' DNA polymerases, *Proc Natl Acad Sci U S A 107*, 20305-20310.

7. Steitz, T. A. (1992) A mechanism for all polymerases, *PNAS 89(20)*, 9579-83.

8. Abad, M. G., Rao, B. S., and Jackman, J. E. (2010) Template-dependent 3'-5' nucleotide addition is a shared feature of tRNAHis guanylyltransferase enzymes from multiple domains of life, *Proceedings of the National Academy of Sciences 107*, 674-679.

9. Jackman, J. E., and Phizicky, E. M. (2006) tRNAHis guanylyltransferase catalyzes a 3'-5' polymerization reaction that is distinct from G-1 addition, *Proc Natl Acad Sci U S A 103*, 8640-

8645.

10. Jackman, J. E., and Phizicky, E. M. (2006) tRNAHis guanylyltransferase adds G-1 to the 5' end of tRNAHis by recognition of the anticodon, one of several features unexpectedly shared with tRNA synthetases, *Rna 12*, 1007-1014.

11. Rao, B. S. , Mohammad, F.,Gray, M. W., and Jackman, J.E. "Absence of a universal element for $tRNA^{His}$ identity in *Acanthamoeba castellanii*." Nucleic Acids Research. (2012): 1-10. Print.

12. Ledoux, S. and Uhlenbeck, O.C..“3’-^{32}P labeling tRNA w. nucleotidyltransferase for assaying aminoacylation and petide bond formation.” Methods. 44.2 (2008): 74-80. Print.

13. Francklyn, C.S., First, E. A.,Perona, J. J., and Hou, Y. "Methods for Kinetic and Thermodynamic Analysis of Aminoacyl-tRNA Synthetases." Methods. 44.2 (2008): 100-118. Print.

Acknowledgments:

To the following funding sources that made this work possible: Arnold and Mabel Beckman Foundation, Ohio State Chemistry and Biochemistry Department, Ohio State College of Engineering, NIH support from Jane Jackman. I would also like to thank Dr. Jane Jackman for her fantastic mentorship and direction, as well as Dr. Venkat Gopalan and Dr. Keith Gooch for serving on my committee and providing me with constructive comments. Finally, I would like to thank the members of the Jackman Lab for their ongoing guidance, support and friendship.